Genetic variation within and between European cattle breeds

Sarah Caroline Blott
B.Sc (London)
M.Sc (Edinburgh)

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
1997

Contents

Chapter 5 Genetic variation within the Hereford breed of cattle

Chapter 6 Discriminating among cattle breeds using genetic markers

Chapter 7 General discussion

List of tables

List of figures

Acknowledgements

Financial support from the Rare Breeds Survival Trust and the Traditional Livestock Foundation, which enabled this project to be carried out, is gratefully acknowledged. I would also like to thank Lawrence Alderson, of the Rare Breeds Survival Trust, for his interest in the project and helpful comments on the work. Particular thanks are due to my supervisors. Thanks to Chris Haley for his support and patient guidance throughout the project, also to John Williams, Peter Keightley and Robin Thompson for their help and support.

The co-operation of the Hereford Cattle Society in organizing the collection of samples from Herefords is much appreciated. In particular, thank you to all the breeders who submitted blood samples from their animals for testing. I am grateful to Olivia Small and Brian McTier of the Roslin Cattle Blood Typing Service for their help with the typing of samples. Thank you to Olivia for answering my questions on cattle blood types, and allowing me to use your computer during the many months I spent in the lab inputting data. Thanks also to Alastair Cameron for his invaluable assistance with Foxpro programming.

I would like to thank Ricardo Pong-Wong for helping me to extend my knowledge of Fortran programming. Also Dave Waddington for discussions on the graphical representation of genetic distances, Philippe Baret for sharing his expertise on estimating linkage disequilibrium and Chris Theobald for a helpful introduction to the problem of breed identification.

I also thank my fellow contemporary students in the Division of Genetics & Biometry, Annemarie, David, and Dimitri, for their friendship and also their companionship during the hard task of writing a thesis. Finally, thank you to all the staff (Neil, Steve, John, Liz et al) and students (Grant, Katrin et al) who, over the last three years, have made the Division such an enjoyable place to work.

List of publications

Blott, S.C., Williams, J.L., Keightley, P.D. & Haley, C.S. (1996) Monte Carlo methods applied to genetic distances between European cattle breeds. *Proceedings of the 25th International Society for Animal Genetics Conference*, Tours, France.

Blott, S.C., Williams, J.L. & Haley, C.S. (1997) Genetic diversity in European cattle breeds. *Proceedings of the British Society for Animal Science Annual Meeting*, Scarborough, UK.

Blott, S.C., Williams, J.L. & Haley, C.S. (1997) Genetic relationships among European cattle breeds. Submitted to *Animal Genetics*.

Blott, S.C., Williams, J.L. & Haley, C.S. (1997) Genetic variation within the Hereford breed of cattle. Submitted to *Animal Genetics*.

Blott, S.C., Williams, J.L. & Haley, C.S. (1998) Genetic variation within the Hereford cattle breed. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production* **28**, 115-118.

Abstract

Effective conservation strategies for livestock populations will require breeds to be characterized on the basis of their genetic attributes. Genetic markers can be used to survey breed relationships and within-breed population structure. They can also be used for the identification of breeds. In this study markers were used to investigate genetic relationships among thirty-seven European cattle breeds. Blood group (A,B,C,F,L,S,Z) and serum protein (transferrin and albumin) polymorphisms in 18,859 animals were analysed. French, Italian and Channel Island breeds were found to have generally higher heterozygosities and greater numbers of alleles than breeds from Britain and Northern Europe. Genetic distances among breeds were estimated, and two major breed groups were identified. French, Italian and Channel Island breeds clustered together in one group, while the second group consisted of breeds from Britain and Northern Europe. All breeds were significantly different from one another ($p < 0.0001$). In general, relationships among breeds reflected their geographical origin and common ancestry, rather than the agricultural use for which the breeds have been selected.

Genetic differentiation among populations of Hereford cattle from different countries, and change in the genetic structure of the British Hereford population over time were investigated using blood typing data collected between 1966-1996. Genetic distances were estimated among Hereford populations from Britain, Ireland, Sweden, Canada and New Zealand, and six other beef breeds. The Hereford populations clustered together, and were significantly different from the other breeds. Irish, New Zealand and Swedish Herefords were significantly different from most other Hereford groups. Canadian Herefords were significantly different from groups with 100% British ancestry, but not from groups with hybrid (British and Canadian) ancestry or polled groups. However, there was no distinction between horned and polled groups from the same country. The proportion of Canadian genes in the hybrid population was estimated at 0.65 (± 0.21). Canadian Hereford groups were found to be least heterozygous of the Hereford groups. The replacement of the British Hereford population with Canadian animals may lead to a loss of genetic variation in the long-term.

Genetic markers provide a potentially powerful way of identifying the breed of individual animals. Microsatellite markers were found to be more powerful than diallelic markers for distinguishing among breeds, approximately one microsatellite to six diallelic markers was required to achieve the same power of discrimination. The number of markers needed to achieve a given error rate could be reduced by selecting the most discriminatory individual markers. However, the exact number of markers required depends on the number and type of breeds to be identified. The development of new DNA technologies is making screening of large numbers of loci economically viable, which will enable genetic differences between breeds to be more precisely understood and their genetic histories to be more accurately defined.

Key to breed abbreviations

Code	Breed	Region of origin
AA	Aberdeen Angus	Scotland
AY	Ayrshire	Scotland
BA	Blonde d'Aquitaine	France
BB	Belgian Blue	Belgium
BG	Belted Galloway	Scotland
BS	Brown Swiss	Switzerland
BU	Butana	North Sudan
BW	British White	England
CH	Charolais	France
CI	Chianina	Italy
DX	Dexter	Ireland
GA	Galloway	Scotland
GB	Gelbvieh	Germany
GL	Gloucester	England
GO	Gobra	Senegal
GU	Guernsey	Channel Islands
HA	Hariana	North India
HF	Hereford	England
HL	Highland	Scotland
HO	Holstein Friesian	Netherlands
IC	Icelandic	Iceland
JS	Jersey	Channel Islands
KE	Kenana	South Sudan
KY	Kerry	Ireland
LM	Limousin	France
LO	Longhorn	England
LR	Lincoln Red	England
MA	Maine-Anjou	France
MC	Marchigiana	Italy
ME	Maure	Mauritania
MR	Meuse Rhine Yssel	France
MU	Murray Grey	Scotland
ND1	N'Dama	Gambia
ND2	N'Dama	Guinea
ND3	N'Dama	Guinea Bissau
ND4	N'Dama	Nigeria
ND5	N'Dama	Senegal
PH	Poll Hereford	England
PM	Piemontese	Italy
RM	Romagnola	Italy
SA	Salers	France
SD	South Devon	England
SH	Shorthorn	England
SM	Simmental	Switzerland
SW	Sahiwal	Pakistan
SX	Sussex	England
TH	Tharparker	North India
WB	Welsh Black	Wales
WF	White Fulani	Nigeria
WP	White Park	England

Hereford populations

B60	British born 1960-69
B70	British born 1970-79
B80	British born 1980-89
CN60	Canadian born 1960-79
CN80	Canadian born 1980-96
CP60	Canadian Poll born 1960-79
CP80	Canadian Poll born 1980-96
H70	Hybrid born 1970-79
H80	Hybrid born 1980-89
H90	Hybrid born 1990-96
IR	Irish
NZ	New Zealand
P60	Poll (British) born 1960-69
P70	Poll (British) born 1970-79
P80	Poll (British) born 1980-89
P90	Poll (British) born 1990-96
SW	Swedish
TR	Traditional British

Chapter 1

Introduction and literature review

1.1 Introduction

Animal populations exhibit diversity for traits such as height, weight, conformation and coat colour, as well as many other characteristics. Genetic variation underlying these traits provides the foundation on which artificial and natural selection acts. Without genetic variation man would not have been able to shape the development of domestic species to meet his requirements, nor would natural selection have led to the adaptation and evolution of species. The field of population genetics is concerned with studying the levels and distribution of genetic variation in populations, and tries to explain the variation in terms of its origin, maintenance and evolutionary importance (Hartl 1981). Analysis of the global distribution of genetic variation within endangered species has important implications for the maintenance of biodiversity, and many studies have been carried out on wildlife populations. More recently there has been a growing concern with the conservation of genetic diversity in domestic livestock species (Food and Agriculture Organization 1996).

Domestic animals have been selectively bred by humans over many centuries. Diverse breeds have arisen, that are adapted to their local environment and selected for production traits of value to the human communities exploiting them. Initially breeds would have been reproductively isolated due to geographic separation and, in the last one to two hundred years, they have been isolated because of barriers imposed by pedigree herd book registration. As a result of this genetic isolation, genetic drift will have occurred. In addition, breeds have undergone natural and artificial selection. Natural selection will have led to breeds becoming adapted to different environments including different climates, food sources and disease challenges. Breeds may also have been artificially selected for different purposes, e.g. beef or milk production, and for different phenotypes such as coat colour. The action of all these forces has resulted

in breeds that are genetically differentiated, and that perhaps have unique genetic adaptations allowing them to exploit different environmental or economic niches. The current interest in genetic diversity of livestock breeds has arisen as a number of breeds have become extinct, and others are declining numerically. Genetic identification and classification of livestock breeds has become a global objective, with the aim of identifying unique genetic resources in order that action can be taken to preserve them. The preservation of genetic resources being essential if man is to be able to maintain flexibility to meet future changes in agricultural requirements (Food and Agriculture Organization 1996).

Individual breeds represent separate and unique gene pools, and the preservation of whole gene pools may be the best strategy that is currently available for preserving the genetic adaptations of breeds (Hall and Bradley 1995). The process of breed development, however, is dynamic rather than static. Over the years commercial animals have been produced by grading up from one breed to another or by forming new synthetic breeds from two or more existing breeds. Modern cattle breeds that have been developed from such crosses include the Belgian Blue (Shorthorn/Black-and-White), Maine-Anjou (Shorthorn/Mancelle), Murray Grey (Shorthorn/Aberdeen Angus) and the Luing (Shorthorn/Highland) (Porter 1991). Introgression of genes from one breed to another has also occurred. Well documented examples in cattle include introgression of the poll gene (absence of horns) into several breeds. For example, the gene was introduced into Hereford cattle from the Red Poll and Galloway breeds (Heath-Agnew 1983), and has also been introgressed into other European breeds, e.g Charolais. The double-muscling gene, which is thought to have originally occurred in the Shorthorn breed (Grobet et al 1997), is now found in both the Belgian Blue (a Belgian breed) and the Asturianas (a Spanish breed). The best known example of breed replacement has been the introduction of the North American Holstein into European black-and-white cattle populations such as the British Friesian. A similar process is also happening in red-and-white cattle populations, with red-and-white Holstein sires being used in Ayrshire populations and also some of the red-and-white Scandinavian breeds. Traditionally particular breeds have been identified by

characteristic phenotypes, for example, the Aberdeen Angus usually has a black coat colour (although red Angus do exist) and is polled. These traditional characteristics may now be changing, as breeders develop new lines. In Canada a new type of animal, with a black coat colour, has been produced in the Simmental breed. The Simmental population originally contained only red-and-white coat colour genes, and the black colour gene was introgressed from the Aberdeen Angus.

Introgression of genes from one breed to another or breed replacement may not be the only threats to genetic diversity. Some pre-eminent cattle breeds, such as the Charolais, Limousin and Hereford among others, have been exported to countries outside their native ones, and are found all around the world. Different national populations of the breed now exist, with each population having been derived from a different founder group and perhaps selected for different objectives. In the last twenty years animals bred in Canada or North America have been imported into European populations, and these bloodlines now predominate in breeds such as the Holstein or the Hereford. This process has been accelerated by the use of artificial insemination (AI) and embryo transfer (ET). However, the global use of a small number of selected sires will result in a sharp decline in the effective population size of a breed (Goddard 1992), and the subsequent loss of genetic variation may have serious implications for the future of the breed.

When introgression or breed replacement takes place valuable genetic resources may be eroded. If breeds are specially adapted to particular environments and have unique gene combinations these may be lost during the process of grading up or replacement. Conservationists are now concerned about understanding the level of diversity and adaptation in different livestock breeds, and assessing the effect of breed introgression and replacement on the genetic variation within populations. This chapter reviews some of the methods that have been developed by population geneticists for assessing the genetic structure of populations, together with application of these methods to studies of genetic diversity in livestock species.

1.2 Genetic variation

At the molecular level genetic variation in populations is introduced by mutation, which is a change in the DNA from the normal, or most common, sequence. Variation exists in most populations in the form of multiple alleles at many individual loci (Hartl 1981), so that individuals differ in the alleles that they possess. Polymorphic loci are defined as loci where the most common allele in the population has a frequency of less than 0.95 (Hartl 1981). The presence of polymorphism in a population is a measure of its genetic variability, and the genetic variation can be quantified by measuring allele frequencies.

1.3 Genetic markers

With the development of electrophoresis in the 1960s it became possible for geneticists to directly measure the variation at individual loci. Electrophoresis was originally developed to detect differences among protein variants, i.e. the final product of the gene. With the advent of DNA technology other markers have become available that directly reflect variation in the DNA, including restriction fragment length polymorphisms (RFLPs), DNA fingerprinting, microsatellite markers, and randomly amplified polymorphic DNA (RAPDs). The use of these markers to measure genetic diversity in both natural and domesticated populations has closely followed the development of each marker type.

Initial studies on livestock breeds utilised blood type or other protein loci to measure diversity in cattle, sheep, pig and goat breeds (Baker and Manwell 1980; Grosclaude et al 1990; Buis and Tucker 1983; Zanotti Casati et al 1990; Van Zeveren et al 1990a, 1990b; Tunon et al 1989). A variety of DNA markers have subsequently been used to study diversity in livestock species. RFLPs have been used to study zebu-taurine variation in Y chromosomal DNA in cattle (Bradley et al 1994), while RAPDs have been used to assess diversity in cattle, chickens and turkeys (Gwakisa et al 1994; Smith et al 1996). DNA fingerprinting has been the method chosen for the analysis of genetic

diversity in poultry (Dunnington et al 1991; Siegel et al 1992; Dunnington et al 1994; Zhu et al 1996) and has also been applied to sheep (Hermans et al 1993).

Comparison of sequences from mitochondrial DNA has proved informative when studying relationships among different species (Avise et al 1987; Moritz et al 1987). Mitochondrial DNA is preferred for sequence based phylogenetic analysis because it occurs as a single copy molecule, which is technically easy to sequence, and it is also not subject to recombination. However, mitochondrial DNA may be less suitable for studying relationships among closely related populations. Loftus et al (1994) have shown that while there is variation in the DNA sequence when the two cattle species, *Bos taurus* and *Bos indicus*, are compared, there is very little variation in mitochondrial DNA sequence among European cattle breeds (*Bos taurus*).

The most informative markers for distinguishing among closely related or intraspecies populations, such as livestock breeds, are thought to be microsatellite markers (Goldstein and Pollock 1997). Microsatellite markers have already been demonstrated to be useful for estimating relationships among cattle (MacHugh et al 1994; Moazami-Goudarzi et al 1994; Ciampolini et al 1995; Basedow et al 1996) and sheep breeds (Buchanan et al 1994). The problem of co-ordinating studies to establish breed relationships using microsatellite markers is being tackled by the Food and Agriculture Organization (FAO) of the United Nations. They have established a project (MoDAD) which is designed to evaluate the relationships among a number of breeds in each of the livestock species based on an agreed set of microsatellite markers. This project is currently in its early stages and no results are available as yet. It is also not planned to sample all breeds (Food and Agriculture Organization 1996), and the cataloguing and conservation of genetic diversity for all breeds within one country or continent will be the responsibility of other regional organizations.

1.4 Factors influencing genetic variation and methods of measuring genetic variation

Genetic variation can be partitioned into two components intra- or within population variation and inter- or between population variation. The following sections (1.4.1 to 1.4.3) describe the factors that influence genetic variation within populations, and outline some measures of within population variation.

1.4.1 Heterozygosity

One measure of intra-population variation that can be estimated using marker information is the average heterozygosity, the frequency of heterozygotes averaged over the loci tested (Falconer 1989). This is equivalent to the proportion of loci at which the average individual is heterozygous or the proportion of heterozygous individuals in the population. Heterozygosity can be expressed as an observed value, based on the genotypes observed, or as an expected value, under Hardy-Weinberg equilibrium, calculated from the observed or estimated allele frequencies. The two will differ if there is non-random mating in the population. However, while the expected value can be calculated for either dominant or co-dominant loci, the observed value can only be obtained from co-dominant loci.

1.4.2 Population size and genetic drift

Genetic variation within a population is influenced by the systematic processes of migration, mutation and selection. These processes are systematic because they change gene frequencies in a way that is predictable both in amount and direction (Falconer 1989). Variation is also influenced by the process of drift, which arises in small populations from the effects of sampling. Drift is predictable in amount but not in direction (Falconer 1989). In each parental generation the genes to be passed to the next generation are a sample. Not all individuals in a generation will have offspring, and even if individuals do reproduce, not all their genes will be passed on, due to Mendelian sampling. As a result the gene frequencies are subject to sampling variation

between successive generations, and the smaller the number of parents the greater is the sampling variation.

This sampling process results in random changes of gene frequency or random genetic drift. Random drift occurring independently in different sub-populations leads to genetic differentiation between the sub-populations. At the same time, genetic variation within each sub-population becomes progressively reduced, and the individuals become more and more alike in genotype. There is increased homozygosity, or loss of heterozygosity, which tends to allow the expression of recessive deleterious alleles resulting in the loss of fertility and viability.

1.4.3 Effective population size

The effective population size gives a measure of the rate of genetic drift and inbreeding in the population. It can be computed or estimated from laboratory or field data, and predicted under a range of circumstances (Caballero 1994). Two types of effective size are recognised, the inbreeding effective size reflects the accumulation of gene correlations within individuals while the variance effective size reflects the effects of gene frequency drift. The effective population size can give a prediction of the impact of management practices on genetic variation. It is also important in plant and animal breeding because its magnitude affects the response to artificial selection and the variance of response (for reviews see Hill 1985, 1986).

Various methods of predicting the effective population size are available, including those accounting for systems of mating between relatives (Caballero and Hill 1992a,b), and for populations undergoing selection (Woolliams et al 1993; Wray and Thompson 1990; Wray et al 1990,1994; Santiago and Caballero 1995). The actual effective population size (N_e) can be computed from inbreeding coefficients obtained from pedigrees. Estimations can be made even when there are individuals with uncertain parentage in the pedigree (Perez-Enciso et al 1992). The above methods apply to closed populations where there is no migration or mutation, the estimation of effective

size in sub-divided populations allowing for migration has been considered by Chesser et al (1993).

The use of genetic markers allows genotypes to be distinguished and genotypic frequencies estimated. If the rate of decay of heterozygosity, between generation t-1 and generation t (H_t/H_{t-1}) can be traced then the formulae $\Delta F = \frac{1}{2Ne}$ and $\frac{Ht}{Ht-1} = 1 - \Delta F$ can be used to obtain an estimate of the effective population size (N_e). Kimura (1983) gives further details on how an estimate of the effective population size can be obtained from the observed heterozygosity. The effective population size can also be estimated from temporal changes in allele frequency (Waples 1989).

Despite the fact that livestock populations, in general, have well-recorded pedigrees, there have been very few published estimates of effective population sizes based on pedigree information. Estimates that have been made have tended to focus on rare or minority breeds (O'hUigin and Cunningham 1990; Boichard et al 1997). Other studies have estimated rates of inbreeding (Miglior et al 1992; Te Braake et al 1994; Miglior and Burnside 1995) or proportion of founder contributions in present day populations (Woolliams and Mantysaari 1995), based on pedigree information. There have been no published studies that have estimated the effective size of livestock populations based on information from genetic markers.

1.5 Genetic structure and between population variation

Random mating means that any individual has an equal chance of mating with any other individual in the population (Falconer 1989). In practice, livestock populations are structured into herds and families. There may also be structuring of the population by herd book section, according to some phenotypic criteria (e.g. whether animal is polled or horned, or has a particular coat pattern such as a white belt), or by geographical distribution. This structuring of the population means that individuals

within one sub-population are less likely to mate with individuals from another sub-population than they are with individuals in the same sub-population.

In a sub-divided population random genetic drift leads to the genetic divergence of sub-populations. Migration or gene flow between sub-populations may either constrain genetic differentiation, by preventing adaptation to local conditions, or promote differentiation by spreading new genes and combinations of genes through a population (Wright 1931). Selection can act to cause genetic divergence or convergence depending on the circumstances. If the environments in which each sub-population exists are very different then natural selection will promote genetic divergence, as each sub-population becomes adapted to its own environment. Conversely, if the environments are similar then natural selection may prevent genetic divergence of the sub-populations.

The genetic structure of any population depends on the balance of evolutionary forces; genetic drift leading to genetic differentiation among sub-populations, selection that may lead to differentiation if selective forces differ among sub-populations or to homogeneity if selective forces are similar, and gene flow among sub-populations leading to genetic homogeneity. Where sub-populations are separated, because of geographic distribution or mating system, the genetic connections among them depend on the amount of gene flow between sub-populations. With extensive gene flow, the population may breed as a single randomly mating unit, while if there is no gene flow then each sub-population is an independent evolutionary unit (Wright's island model (1951)). Where the population distribution is essentially continuous, it can still be genetically structured and show local genetic differentiation if the range of the population is large relative to the distances that individuals move during their lifetime (Wright's isolation by distance model (1951)).

1.5.1 Wright's F statistics

Wright was the first to consider the theoretical analysis of genetic structure, with other models introduced by Kimura (reviewed by Felsenstein 1976). Wright (1951) introduced three parameters F_{ST}, F_{IT}, and F_{IS} (Wright's F statistics) to summarise genetic structure, where:

F_{IS} measures the reduction in heterozygosity of an individual due to non-random mating within its sub-population

F_{ST} measures the reduction in heterozygosity of a sub-population due to random genetic drift

F_{IT} measures the reduction in heterozygosity of an individual relative to the total population. It is the overall inbreeding coefficient of an individual and includes a contribution due to non-random mating within sub-populations (F_{IS}) and another contribution due to the subdivision of the population (F_{ST})

The relationship between these parameters is:

$$(1 - F_{IS})(1 - F_{ST}) = (1 - F_{IT})$$

When allele and heterozygote frequencies are known for sub-populations, these F-statistics may be estimated and used to describe genetic structure. Wright's definition of F-statistics was made in terms of neutral genes and assumed that the number of sub-populations is infinitely large. Nei (1973) introduced another approach to the study of genetic differentiation of populations, allowing for selection, migration and multiple alleles and a finite number of sub-populations. Previous attempts to reformulate the F-statistics in terms of probability (Jacquard 1974) were based on the same assumptions as Wright's, and required knowledge of the pedigrees of individuals within and between sub-populations in the presence of migration.

Nei (1973) showed that the gene diversity of the total population can be partitioned into its component intra-subpopulational and inter-subpopulational gene diversities, when gene diversity is defined as the heterozygosity expected under Hardy-Weinberg equilibrium. His analysis was designed to estimate the average gene diversity for a large number of loci among a finite number of sub-populations, whereas Wright's F-statistics were intended to be applied to a single locus. However, Nei's gene diversity analysis can also be applied to a single locus. The gene diversity is defined by using the gene frequencies in the present generation, so that no assumption is required about the pedigrees of individuals, selection, and migration in the past.

Nei (1977) further defined the F-statistics in terms of expected and observed heterozygosities, and modified these definitions to allow for sampling of gene frequencies from finite populations (Nei and Chesser 1983). Nei and Chakravarti (1977) also studied the evolutionary change of F_{ST} in a subdivided population of finite size with and without migration. Chesser (1991) considered the influence of gene flow and breeding tactics on gene diversity within populations. Expressions were derived describing the accumulation of gene correlations within and among lineages and individuals, under a model permitting different migration rates by males and females and accounting for various breeding tactics within lineages. The degree of genetic differentiation among breeding groups, inbreeding coefficients, and gene correlations within lineages were found to be primarily functions of breeding tactics within groups rather than gene flow among groups.

F_{ST} is generally estimated as the standardized variance of gene frequency in empirical studies. For a particular locus with two alleles, if p is the mean allele frequency over several sub-populations and σ^2 the variance of the allele frequency over the same sub-populations, then:

$$F_{ST} = \sigma^2/[p(1-p)]$$

When there are more than two alleles per locus and data for more than one locus, there are a variety of ways of combining information to yield a single estimate of F_{ST} (see Weir and Cockerham 1984).

1.5.2 Genetic distance

An alternative measure of population differentiation is provided by the genetic distance between populations, which also allows information from several loci to be combined into a single statistic. Genetic distance can be thought of as a single number expression of how much difference there is in genetic constitution between two populations (Smith 1977). If two populations have precisely the same gene frequencies they will be considered to be at zero distance apart, and the greater the difference between the gene frequencies the greater the distance between populations.

One reason for considering genetic distance is simply to reduce a complicated mass of data to a more manageable form. Two populations can be described by their gene frequencies at several loci, genetic distance reduces this list of frequencies to a single figure representing the overall difference between the populations. The second reason is to try to reconstruct something of the evolutionary history of the populations concerned. Two populations may be similar, with small genetic distance apart, if they are recently descended from a common ancestral population. However, there might be other reasons for similarity, for example, selection could have brought gene frequencies closer together, or one population could be descended from the other, or there could have been introgression from a third population into both. If a set of populations cluster together, or have similar genetic constitutions, it is possible that they share a recent common ancestry, whereas if they are a large distance apart they may have been separated for a long time.

Many different measures of genetic distance have been derived, each with slightly different properties and based on slightly different evolutionary models. The question of which measure is the best to use depends on the objectives or goals of the study in

hand (Swofford and Olsen 1990). Descriptions of some of the more commonly used genetic distances are given below. All the distance measures described are appropriate for transforming allele frequency data into genetic distances.

1.5.3 Nei's standard genetic distance

One of the most frequently used genetic distances has been that of Nei (1972,1978). Nei's (1972) standard genetic distance is defined as:

$$D_N = \ln\left(\frac{J_{xy}}{\sqrt{J_x J_y}}\right)$$

where J_x, J_y and J_{xy} are the arithmetic means across loci of Σx_i^2, Σy_i^2 and $\Sigma x_i y_i$ respectively (where x_i and y_i are the frequencies of the ith allele at a particular locus in populations X and Y). This equation gives a biased estimate when sample sizes are small; an unbiased estimate of the standard distance is obtained by replacing Σx_i^2 and Σy_i^2 with $(2n\Sigma x_i^2 - 1)/(2n_x - 1)$ and $(2n\Sigma y_i^2 - 1)/(2n_y - 1)$ (Nei 1978). Nei's distance assumes an infinite alleles model of mutation, in which there is a rate of neutral mutation and each mutation is to a completely new allele. It is also assumed that the genetic variability initially in the population is at equilibrium between mutation and genetic drift, with the effective size of each population remaining constant.

1.5.4 Rogers' genetic distance

Another widely used distance measure is that of Rogers (1972):

$$D_R = \frac{1}{L}\sum_{1}^{L}\frac{1}{2}\sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}$$

where L = number of loci, and N = number of alleles.

It is simply the Euclidean distance, i.e. squared difference, between the allele frequency vectors for each locus of the two populations being compared, multiplied by a scaling factor. Both Roger's and Nei's distances have the limitation that they are heavily influenced by within-population heterozygosity (Wright 1978; Hillis 1984). Thus, the distance between two populations that are fixed for alternate alleles exceeds that between two populations in which one or both are heteroallelic but have no alleles in common.

1.5.5 Cavalli-Sforza and Edwards' genetic distance

An alternative Euclidean measure, that overcomes the limitation of Rogers' and Nei's distances, is the arc distance of Cavalli-Sforza and Edwards (1967) which is given by:

$$D_{arc} = \sqrt{\frac{1}{L}\sum_{1}^{L}\left(\frac{2\theta}{\pi}\right)^2}$$

where $\theta = \cos^{-1}\Sigma(x_i y_i)^{1/2}$. If no alleles are shared between a pair of populations, the distance takes its limiting value of one regardless of the variability within either population. The distance is standardized with respect to random drift, so that the rate of increase in genetic distance under drift is nearly independent of the initial gene frequencies. The Cavalli-Sforza and Edwards (1967) arc distance and its relative, the chord distance, thus incorporate realistic assumptions about the nature of evolutionary change in gene frequencies without the undesirable properties of the Nei (1972, 1978) and Rogers (1972) measures.

1.5.6 Genetic distance of Reynolds et al

Reynolds et al (1983) have derived a genetic distance based on the coancestry coefficient (θ). The coefficient, θ , is the probability that a random pair of genes at the same locus within a randomly chosen population are identical by descent.

It is related to the time of divergence *t* in the following way:

$$\theta = 1 - (1 - 1/2N_e)^t$$

For short-term evolution, i.e. when t/N_e is small, θ has an approximately linear relationship to time

$$\theta = t/2N_e$$

or

$$D = -\ln (1 - \theta) = t/2N_e$$

Both Cavalli-Sforza and Edward's (1967) distance and Reynolds et al (1983) distance assume that there is no mutation, and that all gene frequency changes are by genetic drift alone. However, they do not assume that population sizes have remained constant and equal in all populations in the way that Nei's distance does. They have expectations that rise linearly with the sum over time of $1/N_e$, where N_e is the effective population size. Thus if population size doubles, genetic drift will be taking place more slowly, and the genetic distance will rise only half as fast with respect to time.

The genetic distance measures described above are all based on the infinite allele model (IAM) of mutation, which is appropriate for protein or biochemical loci, i.e. all mutations are to completely new alleles, with no limit on the number of alleles that can exist. This model of mutation, however, is not appropriate for microsatellite markers, which are thought to follow a stepwise mutation model (SMM) with new alleles being formed by the addition or subtraction of base pair repeats. To allow for this difference in mutation model new distances, appropriate to microsatellite markers, have been derived (Goldstein et al 1995; Shriver et al 1995; Slatkin 1995).

1.5.7 Phylogenetic trees to represent breed relationships

Phylogenetic trees are used to represent evolutionary relationships among species, and may be constructed from gene sequence or haplotype information, or the pairwise genetic distances between species. Methods for constructing phylogenies have been reviewed by Nei (1987), Swofford and Olsen (1990), and Li and Graur (1991). Phylogenetic trees may be rooted or unrooted, with a rooted tree indicating a temporal ordering of species on the tree while an unrooted tree reflects only the distances among species with no indication of which species were ancestral to which (Weir 1996). Methods which construct phylogenies from distance matrices include the UPGMA (unweighted pair group method using an arithmetic average) (Sneath and Sokal 1973), and the neighbour joining method (Saitou and Nei 1987). Maximum parsimony and maximum likelihood methods are applicable to character data, such as sequence information (Weir 1996).

Numerous studies have estimated genetic distances among livestock breeds, and illustrated the relationships among them using a phylogenetic tree (Baker and Manwell 1980; Grosclaude et al 1990; Zanotti Casati et al 1990; Buchanan et al 1994; MacHugh et al 1994; Medjugorac et al 1994). However, it may not be strictly appropriate to represent livestock populations in this way. The linear bifurcating structure of a phylogenetic tree, at least when interpreted as an evolutionary representation of the populations, assumes that the populations split from an ancestral population, and have been subsequently isolated (Hall and Bradley 1995). As previously described, the histories of livestock breeds do not follow this model and contain many examples of populations being crossed or introgression occurring. Nevertheless, phylogenetic trees may provide useful summaries of the relationships among breeds (Takezaki and Nei 1996), particularly if the tree is unrooted and makes no assumptions about temporal relationships among the breeds.

1.6 Estimating gene flow between populations

The amount of gene flow that occurs between populations can be estimated by direct methods, for example using livestock pedigree information, or by indirect methods using allele frequencies or DNA sequences (Slatkin 1987). Three indirect methods are available for estimating the levels of gene flow among populations. The first method is Wright's statistic F_{ST}. Wright showed that for neutral alleles, F_{ST} is approximately equal to $1/(1+4Nm)$, where N is the local population size and m is the average rate of immigration in an 'island' model of population structure. The island model assumes that every local population is equally accessible from every other. One reason for estimating Nm is that this combination of parameters indicates the relative strengths of gene flow and genetic drift. Genetic drift will result in substantial local differentiation if $Nm < 1$ but not if $Nm > 1$ (Wright 1931). If an independent estimate of N is available, from pedigree or census data, then m can be inferred.

The second method for estimating Nm is Slatkin's (1985) method using data for rare alleles (alleles found in only one or a few local populations) which does not depend on the assumption of neutral loci. The method is based on the fact that the average frequency of alleles found in only a single population is a function of Nm:

$$\ln[p(1)] = a\ln(Nm) + b$$

where $p(1)$ is the average frequency of alleles found in only one population sampled and a and b are constants that depend on the number of individuals sampled per population. Barton and Slatkin (1986) found the two methods to be different ways of estimating the same essential properties of gene frequency distributions, and simulation studies found estimates using these two methods to be consistent over a wide range of assumptions about population structure, selection and mutation.

The third method was developed for assessing gene flow among human populations. The gene flow has been estimated directly by studying the changes in gene frequencies,

which result from admixture of previously isolated genetically different populations (Long 1991; Chakraborty et al 1992; Williams et al 1992). This approach is made possible by the availability of information on the demographic history of human populations. The estimate of admixture in the hybrid population is based on the allele frequencies observed in the parental populations (where the frequencies in the parental populations are estimated from modern-day groups believed to represent the parental populations).

Allele frequencies in the admixed or hybrid population can be predicted according to the proportion of each parental population in the population (Long 1991). If P_{hi}, P_{1i} and P_{2i} represent the frequency of allele i in the hybrid population, the first parental population, and the second parental population respectively then:

$$P_{hi} = \mu P_{1i} + (1-\mu) P_{2i},$$

where μ is the proportional contribution of the first parental population, and 1 - μ is the proportional contribution of the second parental population. A weighted least squares estimate of μ can then be obtained (Long 1991) by rewriting the above equation as:

$$(P_{hi} - P_{2i}) = \mu(P_{1i} - P_{2i}) + \varepsilon_i,$$

where ε_i is the error due to sampling and genetic drift. Long and Smouse (1983) used an iterative method of obtaining the weighted least square estimate of admixture, while Chakraborty et al (1992) have derived the closed form expression of the weighted least square estimate.

The method described by Chakraborty et al (1992) has been utilised by MacHugh et al (1997) to estimate the proportion of zebu admixture in nine African populations of cattle. They used Indian zebu, to represent a putative Asian zebu parental population, and Guinean N'Dama, to represent the putative African taurine parental population.

Estimates of the proportion of Zebu in the nine populations ranged between 83.2 % in the Kenana breed, from South Sudan, to 6.9 % in the N'Dama, from Guinea Bissau. Standard errors of the estimates ranged between 1.3 to 2.6 %.

The methods described above cannot conclusively prove that gene flow or admixture has taken place between populations. Selection and drift will also have influenced the gene frequencies within populations, and it is not easy to disentangle their effects from the effect of admixture or hybridization. Realistic predictions about the levels of gene flow among populations perhaps can only be made when demographic information about the history of the populations is available.

1.7 Objectives of the thesis

This thesis focused on the use of genetic markers to evaluate genetic relationships, and levels of genetic differentiation, among British and European cattle breeds. The markers used in the analyses were, predominantly, red blood cell antigen and serum protein loci. Chapters two and three explore different ways of treating the data, when estimating genetic distance and its standard error, that allow the linkage between loci to be taken into account. The genetic relationships among nineteen British and eighteen other European cattle breeds are investigated in chapter four. Chapter five presents a detailed study of the Hereford cattle breed. The level of genetic differentiation among populations of the Hereford from different countries is assessed, and change over time in the genetic structure of the British Hereford population investigated. The effect of interbreeding between the British and Canadian Hereford populations is also considered. The possibility of using genetic markers in a test of breed identity is investigated in chapter six. Two types of marker, diallelic and microsatellite, are compared using simulated data based on observed allele frequencies. The number of markers required to achieve given error rates for discriminating among pure and crossbreed European cattle populations are assessed.

Chapter 2

Genetic analysis of cattle blood types

2.1 Introduction

The data analysed in this thesis were provided by the Roslin Cattle Blood Typing Service. The Service was founded in 1966, to provide cattle parentage testing and identification based on serological tests. The primary reasons for animals being blood typed were as part of the procedure for herd book registration (e.g. random typing of 1 in every 500 registrations is carried out by some breed societies) or for identification of animals being exported. In addition typing was carried out on bulls to be used for AI or that were performance tested, and for the identification of embryo transfer (ET) donors and calves. Around 5000-6000 samples per year have been typed, providing data on a range of different breeds.

2.1.1 Genetic basis of cattle blood types

Cattle red blood cells possess surface antigens that are genetically determined and highly polymorphic. In blood typing, each of the antigenic factors is detected by the binding of a specific antibody (reagent) which, when rabbit complement is added, causes lysis of the red cells. Related antigenic factors are grouped into systems, and eleven red cell systems have been identified: A, B, C, F, J, L, M, S, Z, R', and T'. The systems are independently inherited (i.e. the genes controlling them are on different chromosomes) (Eggen and Fries 1995), however, genes coding for the antigenic factors within the same system are tightly linked (Bouw et al 1974; Grosclaude et al 1983). In the following descriptions of cattle blood types the word *factor* is used to mean a red cell surface antigen that is detected using a single antibody, the expression of the antigen being believed to be controlled by a single dominant locus (Neimann-Sorensen 1956). Factor is thus used as a synonym for locus, with expression of the antigenic factor determined by an animal's genotype at this locus. Animals that are homozygous

or heterozygous for the dominant allele express the antigenic factor on the surface of their red blood cells, while those that are homozygous for the recessive allele do not express the antigen (they have a 'null' phenotype). The word *allele* is used to describe a gene underlying the expression of the red cell antigen. As described above, most of the red cell antigens or factors are regarded as dominant loci having two alleles coding for 'presence' or 'absence' of the antigenic factor. However, one red cell antigen system, the F system, has two antigenic factors that are allelic to one another (Rendel 1967), and in this case the two antigenic factors are described as alleles.

A total of 104 antigenic factors have been identified in the eleven systems. Not all factors have been tested for routinely by the typing laboratories, the antibodies or reagents used in a laboratory may vary over time and will also vary between laboratories. Reagents used by the Roslin Cattle Blood Typing Service over the last thirty years, recorded at three monthly intervals, are shown in Appendix I. A subset of blood type systems (and antigenic factors within those systems) and serum proteins that have remained relatively constant in the panel of loci used by the Service were identified. These were the red cell systems A, B, C, F, L, S, Z and the serum proteins transferrin and albumin. Since the loci controlling the seven blood type systems and two protein loci are located on different chromosomes this provides markers on nine out of the thirty chromosome pairs in cattle. Details of the chromosomal location of the blood type system or serum protein loci, and their individual antigenic factors (dominant loci) or alleles (co-dominant loci) are shown in table 2.1.

Six of the blood type systems consist of dominant loci. The L and Z systems each have a single antigenic factor. The A system has four recognised factors, but only the factor A was included here. The B system is the most polymorphic, with more than thirty factors of which eighteen were included in this study. The second most complex system, after the B, is the C system with seven recognised factors, of which four were included in the study. Three S system factors (S, U and H'), out of the seven recognised factors in this system, were included in this study. One blood type system, the F system, consists of two factors, F and V, which are allelic to one another i.e. the

locus is co-dominant. The two serum protein loci, transferrin and albumin, are both co-dominant. Several transferrin alleles have been identified, but only A, D1, D2 and E are commonly found in European cattle populations, and the albumin locus has two alleles A and B.

Table 2.1 Cattle blood type systems and serum proteins with their constituent antigenic factors (dominant loci) or alleles (co-dominant loci)

Locus	Chromosomal location	Length of coding region	Factors or alleles
Dominant loci			*Factors*
Red cell antigen A	15	-	A
Red cell antigen B	12	0.7 cM	B G I K O P Q T Y A' D' E' G' I' K' O' P' Q'
Red cell antigen C	18	0.3 cM	C R W X
Red cell antigen L	3	-	L
Red cell antigen S	21	-	S U H'
Red cell antigen Z	10	-	Z
Co-dominant loci			*Alleles*
Red cell antigen F	Unknown	-	F V
Transferrin	1	-	A D1 D2 E
Albumin	6	-	A B

Some of the antigenic factors consist of subtypes (Stormont 1955), which are recognised by related antibodies, for example, the antigen Y (from the B system) has two subtypes Y_1 and Y_2. Animals that are phenotypically Y_1 express both the subtype antigens Y_1 and Y_2 (and can have offspring with either of these subtypes), but Y_2 animals express only the Y_2 subtype antigen. Each subtype is identified by the use of specific antibodies, and in most laboratories not all the antibodies required to determine

each subtype will be used. If the Y_1 antibody is used, then animals positive for the Y_2 subtype antigen will be detected, but if only the Y_2 antibody is used, then Y_1 animals will not be detected. Since the subtype reagents used by the Roslin Cattle Blood Typing Service have changed over time (see Appendix I) it was decided to group the subtypes and score animals as either positive or negative for the antigenic factor. For example, one animal may be Y_1 positive and another Y_2 positive, but both would be scored as Y positive.

It has been known for a long time that the blood type factors in the multi-factorial systems (B, C and S systems) are inherited together as a 'phenogroup' (Stormont 1955), the concept being identical to that of a 'haplotype' (Grosclaude et al 1979). An example of a typical phenotype for an individual animal at the B blood type system might be $BO_1Y_1A'D'E'_3$ (by convention only the factors for which the animal tested positive are listed). Information on the blood types of the parents of the animal, or its offspring, would allow its phenotype to be broken down into a genotype consisting of two groups of factors $BO_3Y_1A'E'_3/BO_1Y_1D'$. As described above, for the purposes of this study the antigenic subtypes were grouped. The phenotype and genotype of this animal would then have been scored as BOYA'D'E' and BOYA'E'/BOYD'.

A tentative linkage map of the B system, has been put forward by Grosclaude et al (1979) and Ruiterkamp et al (1977), and for the C system by Guerin et al (1981). It is postulated that the genes controlling the antigenic factors of these systems form a tightly linked cluster of loci, for which a linear order of genes can be determined.

2.1.2 Genetic analysis of blood types

A principal aim of this thesis was to assess differences among cattle groups by estimating allele frequencies within the groups, and comparing genetic distances between them. The complex blood type systems posed a problem, because there is linkage and presumably linkage disequilibrium between the component loci, and measures of genetic distance assume linkage equilibrium between loci, i.e. that allele

frequencies at these loci are independent. One solution to the problem of non-independence of loci is to combine the genotypes at linked loci, and look at the haplotypes formed. If the recombination rate between the component loci is assumed to be effectively zero then the haplotypes may be used as alleles of a compound locus (Ott 1991). Family information is needed to identify the segregating haplotypes in a particular population. However, given the complexity of, for example, the cattle B system which consists up to 30 loci, it is very difficult to identify all haplotypes present in all breeds. Blood group haplotypes in cattle tend to be breed specific (Rendel 1967), and more than 5,000 haplotypes of the B system were observed in data from the thirty-seven European breeds analysed for this thesis.

As it is difficult to identify all the haplotypes segregating in all breeds, previous studies using cattle blood types have simplified the situation by looking at only a limited number of haplotypes (Neimann-Sorensen 1956; Rendel 1967; Grosclaude et al 1990; Andersson-Eklund 1993; Mejdell et al 1993). The most common haplotypes were counted, with rarer haplotypes being grouped into one category. However, grouping of some haplotypes into a single category may lead to the loss of information and the generation of a spurious similarity between breeds. Of further concern would be the loss of phylogenetic information (breed relationships) as a result of recombination between loci resulting in the formation of new haplotypes. These new haplotypes would be classed as completely new alleles, and information on any relationship with existing alleles would be lost. Since recombination between loci will very likely occur at a higher rate than mutation within loci, there is a case for analysing individual loci (antigenic factors) separately rather than treating them as haplotypes. This second type of approach, to treat the factors as independent diallelic loci although they are known to be linked, gives results that are highly correlated with the results obtained using a haplotype approach. Figure 2.1 shows the relationship between the two different methods when applied to data on several Hereford populations (see chapter 5 for details of the Hereford data). The correlation between the two estimates of genetic distance was 0.856, which was found to be significantly different from zero ($p < 0.0001$) using Mantel's test (Mantel 1967, see section 2.2.1 for details).

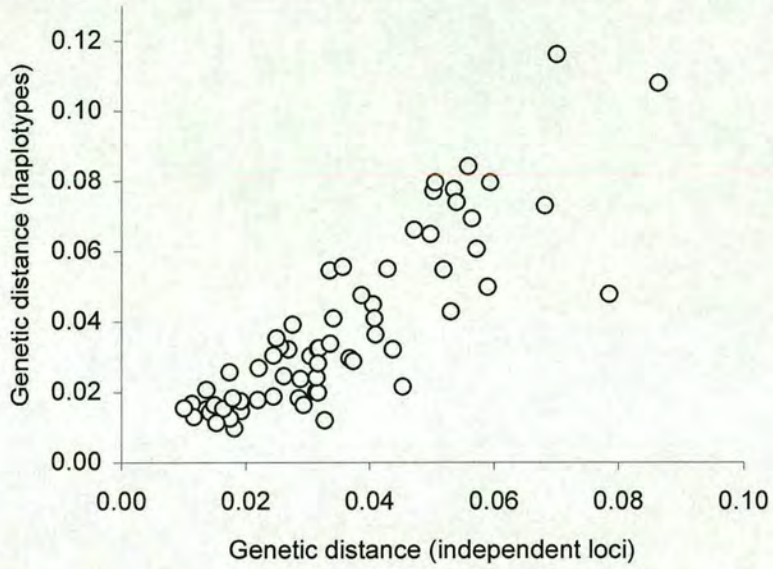

Figure 2.1. Relationship between genetic distances (among 12 Hereford populations) estimated from the B blood type system by two different methods: 1) treating antigenic factors as independent loci or 2) combining factors into haplotypes

Treating the loci as independent may have the advantage that, since the mutation rate at individual loci is lower than the recombination rate between loci, there is likely to be less loss of phylogenetic information. However, since the loci of the blood type systems are very closely linked, it is conceded that there may be some redundancy of information when the allele frequencies at all loci are used to estimate genetic distance. A similar approach to using haplotypes, but which simplified the problem of having to identify a large number of haplotypes, was developed by Grosclaude et al (1990) for their study on genetic relationships among French cattle breeds. They identified five factors within the B system which were only ever seen alone or as pairs together, i.e. they behaved in an almost allelic fashion, or there was strong linkage disequilibrium between these factors. Using the method of Grosclaude et al (1990) the B system can be re-defined as a single locus with six alleles (G_1, I_1, K, P_1, T and a null allele). Any other B system factors occurring with these five factors are then ignored for the purposes of the analysis.

In this chapter the differences between two alternative approaches of utilising the complex blood type systems in genetic distance studies, namely the approach developed by Grosclaude et al (1990) and that of treating the factors as independent diallelic loci, were investigated. In addition the gene frequency correlations between B system factors were estimated, and the patterns of correlations observed in different breeds compared.

2.2 Comparison of two methods of analysing the cattle B blood type system

2.2.1. Methods

Two methods of using information from the B blood type system to estimate genetic distance between breeds were compared. Method one treated the eighteen component factors (see table 2.1) of the B system as independent diallelic loci (referred to as the "independent loci" method). The second method (Grosclaude et al 1990) grouped the B system phenotypes into a limited number of categories based on five factors (G_1, I_1, K, P_1, and T), which were considered to be independent (referred to as the "reduced haplotype" method). The "reduced haplotype" method allows any redundancy of information from linked loci to be taken into account, by re-defining the B system as a single locus with six alleles.

Phenotype counts were made in thirty-seven European cattle breeds, with the B system defined as i) eighteen independent loci (where the loci are the factors as listed in table 2.1), the "independent loci" method, and as ii) a single locus with six alleles (factors G_1, I_1, K, P_1, and T), the "reduced haplotype" method. Allele frequencies were estimated from the phenotype counts, using an iterative procedure (Weir 1996), and genetic distances (Reynolds et al 1983) between the thirty-seven breeds calculated from the allele frequencies. The Pearson correlation coefficient between the two distance matrices was calculated, and its significance tested using Mantel's test (Mantel 1967; Manly 1986).

Mantel's test is a way of assessing whether the elements of two matrices **X** and **Y**, of order L x L, show correlation. A test statistic

$$Z = \sum_i \sum_j x_{ij} y_{ij} \qquad [2.1]$$

is calculated, where x_{ij} and y_{ij} are the corresponding elements of the matrices **X** and **Y** respectively. This test statistic is then compared with the distribution of Z that is obtained when one of the matrices has the order of its rows (or columns) randomized. The Pearson correlation coefficient between the elements of X and Y is given by (Mantel 1967):

$$r = \{Z - E(Z)\} / \left[\left[B_X - G_X{}^2 / \{L(L-1)\} \right] \left[B_Y - G_Y{}^2 / \{L(L-1)\} \right] \right]^{1/2} \qquad [2.2]$$

where $B_X = \sum_{i=1}^{L} \sum_{j=1}^{L} x_{ij}{}^2$, $B_y = \sum_{i=1}^{L} \sum_{j=1}^{L} y_{ij}{}^2$, $G_X{}^2 = \left(\sum_{i=1}^{L} \sum_{j=1}^{L} x_{ij} \right)^4$, $G_y{}^2 = \left(\sum_{i=1}^{L} \sum_{j=1}^{L} y_{ij} \right)^4$

and $E(Z) = \left(\sum_{i=1}^{L} \sum_{j=1}^{L} x_{ij} \right) \left(\sum_{i=1}^{L} \sum_{j=1}^{L} y_{ij} \right) \Big/ \{L(L-1)\}$

The distribution of the Z statistic was obtained by permuting the second matrix 10,000 times. Multidimensional scaling (described in chapter 4, section 4.2.7) was carried out on both distance matrices, and the resulting plots were aligned using Procrustes rotation (Genstat 1993). Procrustes rotation is a procedure by which two configurations of points, in this case points representing the genetic relationships among breeds, can be related to one another. Although the same method, multidimensional scaling, was used to produce the configuration of points it would not necessarily result in configurations that are related to the same axes. Procrustes rotation shifts or rotates the coordinate system for one set of points so that they match, as closely as possible, the coordinates of the other set of points.

2.2.2 Results

The phenotype counts for the B system using the "reduced haplotype" method (Grosclaude et al 1990) are shown in table 2.2. The phenotypic counts observed in the thirty-seven European breeds analysed in this study confirmed the observation of Grosclaude et al (1990), that the antigenic factors G_1, I_1, K, P_1, and T are rarely seen together in combinations of more than two. However, it also notable that a large proportion of animals are allocated to the null category, i.e. have no expression of these five factors. In reality, these animals differ in the expression of other antigenic factors in the B system but under this method of analysis are grouped into one category. The percentage of animals in this null category is smaller for the French and Italian breeds (22-60 %) than for British breeds (62-97 %).

Genetic distances between the thirty-seven breeds calculated using both the "independent loci" and "reduced haplotype" methods are shown in table 2.3. The relationship between the two measures of genetic distance is illustrated in figure 2.2. The correlation between the distance matrices was $r = 0.512$, the observed Z statistic was 25.12. Over the ten thousand permutations used to generate the null distribution for the Z statistic, no values greater than 25.12 were found, hence $p < 0.0001$, and the correlation between the matrices is significantly different from zero. However, figure 2.2 indicates that the "reduced haplotype" method results in some genetic distances that are close to zero compared with the "independent loci" method, particularly when the distance is between two British breeds. In contrast, when the distance is between French and British breeds the "reduced haplotype" method gives some distances that are higher than those obtained with the "independent loci" method. The closest relationship between the two methods exists for genetic distances between two French breeds. These results reflect the effect of grouping animals into the null category (where "null" signifies expression of no antigenic factors), since the five factors that are identified appear to be at higher frequency in the French breeds compared with British breeds. The "reduced haplotype" method is better at discriminating among French than among British breeds. When French and British breeds are compared, distances may

Table 2.2 Phenotype counts for the B system in different breeds with the B system defined as a single locus with six alleles according to the "reduced haplotype" method of Grosclaude et al (1990).

	Breed																		
B system phenotype	AA	AY	BB	BG	BA	BW	BS	CH	CI	DX	GA	GB	GL	GU	HF	HL	HO	IC	JS
G_1			10			1			1	1		1					23		
I_1	4		57		36		4	5	1	5		11	61	42	1		363	5	
K			37	1		2	4	19			3	14	22	8	5		49	5	13
P_1	7	1	9		1		4	340		2		1		4	24		188	8	2
T	13	7	2	14	80	3	26	221	44	84	4	2	6	31	1	13	24		162
G_1I_1			1									1					4		
G_1K												1						1	
G_1P_1																			
G_1T																			
I_1K												1							
I_1P_1								2									7		
I_1T					7		2		2	1			6	10			2		
KP_1								3											
KT							3	4					1	6					54
P_1T							1	40		2	1			1			3		
G_1I_1K												1							
G_1I_1T																			
I_1KT					1				1										
KP_1T								1											
Null	173	146	1126	58	179	65	24	934	17	152	56	66	62	120	1003	44	7115	63	229
Total	197	154	1242	73	304	71	68	1569	66	247	64	99	158	222	1034	57	7778	82	460

Table 2.2 Phenotype counts for the B system in different breeds with the B system defined as a single locus with six alleles according to the "reduced haplotype" method of Grosclaude et al (1990) – continued.

	Breed																	
B system phenotype	KY	LM	LO	LR	MA	MC	MR	MU	PH	PM	RM	SA	SD	SH	SM	SX	WB	WP
G_1				8	2		3								1		6	10
I_1	1	35			1		7	3		3	1	21	6	1	100		5	
K		19			1	1			6		3		1	1	17			1
P_1	2	29	5		19	2			23	4	3			2	11	2	4	
T		626			11	44		5		21	35	15	18		87	5	2	1
G_1I_1					1												1	
G_1K																		
G_1P_1																		1
G_1T						3											1	2
I_1K																		
I_1P_1					2										1		2	
I_1T		17				1		1		1	3	3			5			
KP_1															1			
KT		5													1			
P_1T		13			1	9				2	8	1						
G_1I_1K																		
G_1I_1T																		
I_1KT																		
KP_1T																		
Null	49	1069	61	53	48	17	89	52	1050	38	25	45	59	60	704	57	44	31
Total	52	1813	66	61	86	77	99	61	1079	69	78	85	84	64	928	64	65	46

Table 2.3 Genetic distances between 37 breeds estimated from the B blood group system ("independent loci" method below the diagonal, "reduced haplotype" method above the diagonal).

	AA	AY	BB	BG	BA	BW	BS	CH	CI	DX	HO	GA
AA		0.020	0.007	0.023	0.124	0.008	0.275	0.111	0.386	0.118	0.008	0.002
AY	0.075		0.011	0.072	0.201	0.005	0.358	0.188	0.469	0.195	0.009	0.019
BB	0.048	0.044		0.049	0.161	0.004	0.318	0.149	0.430	0.159	0.001	0.008
BG	0.043	0.067	0.050		0.056	0.049	0.186	0.055	0.286	0.047	0.053	0.023
BA	0.096	0.070	0.062	0.088		0.170	0.053	0.016	0.124	0.004	0.168	0.129
BW	0.117	0.102	0.112	0.057	0.158		0.326	0.158	0.438	0.164	0.004	0.006
BS	0.122	0.176	0.110	0.173	0.132	0.277		0.071	0.022	0.063	0.324	0.280
CH	0.100	0.066	0.067	0.059	0.064	0.108	0.168		0.156	0.018	0.154	0.119
CI	0.224	0.166	0.203	0.212	0.144	0.296	0.206	0.201		0.132	0.437	0.390
DX	0.062	0.074	0.051	0.057	0.071	0.098	0.099	0.050	0.193		0.165	0.121
HO	0.024	0.057	0.020	0.039	0.070	0.097	0.125	0.085	0.224	0.057		0.010
GA	0.032	0.093	0.054	0.039	0.106	0.093	0.169	0.084	0.277	0.059	0.035	
GB	0.066	0.085	0.074	0.050	0.091	0.083	0.182	0.086	0.227	0.077	0.050	0.068
GL	0.154	0.125	0.119	0.160	0.085	0.233	0.159	0.144	0.154	0.143	0.112	0.172
GU	0.086	0.099	0.064	0.086	0.064	0.147	0.140	0.079	0.173	0.072	0.075	0.073
HF	0.148	0.110	0.150	0.172	0.114	0.219	0.249	0.128	0.324	0.122	0.141	0.142
PH	0.152	0.115	0.156	0.174	0.122	0.220	0.261	0.129	0.333	0.124	0.147	0.140
HL	0.106	0.089	0.105	0.099	0.093	0.157	0.206	0.087	0.292	0.092	0.080	0.096
IC	0.060	0.120	0.094	0.065	0.123	0.091	0.209	0.098	0.271	0.091	0.073	0.063
JS	0.228	0.182	0.140	0.214	0.107	0.287	0.158	0.162	0.207	0.152	0.169	0.215
KY	0.092	0.087	0.080	0.143	0.094	0.207	0.104	0.147	0.180	0.106	0.091	0.160
LM	0.107	0.104	0.066	0.084	0.045	0.156	0.131	0.058	0.170	0.059	0.097	0.094
LO	0.104	0.175	0.121	0.179	0.184	0.269	0.149	0.198	0.351	0.129	0.101	0.116
LR	0.172	0.179	0.145	0.193	0.149	0.303	0.117	0.191	0.189	0.154	0.169	0.242
MA	0.086	0.032	0.050	0.080	0.068	0.129	0.128	0.064	0.103	0.072	0.073	0.120
MC	0.193	0.153	0.179	0.174	0.110	0.250	0.179	0.147	0.045	0.145	0.197	0.237
MR	0.061	0.022	0.047	0.053	0.056	0.095	0.181	0.058	0.193	0.077	0.051	0.077
MU	0.041	0.082	0.074	0.053	0.097	0.084	0.155	0.072	0.241	0.038	0.051	0.043
PM	0.051	0.075	0.052	0.042	0.044	0.102	0.127	0.064	0.164	0.050	0.032	0.048
RM	0.134	0.157	0.112	0.129	0.087	0.221	0.099	0.125	0.118	0.099	0.137	0.155
SA	0.098	0.105	0.085	0.090	0.067	0.134	0.207	0.069	0.237	0.082	0.089	0.071
SH	0.088	0.081	0.062	0.089	0.067	0.160	0.090	0.062	0.185	0.048	0.080	0.114
SM	0.063	0.052	0.057	0.028	0.070	0.054	0.197	0.037	0.215	0.056	0.050	0.042
SD	0.031	0.058	0.019	0.028	0.058	0.090	0.120	0.070	0.187	0.045	0.017	0.036
SX	0.055	0.116	0.109	0.096	0.139	0.130	0.169	0.119	0.280	0.055	0.086	0.053
WB	0.048	0.026	0.029	0.035	0.063	0.081	0.141	0.052	0.168	0.052	0.027	0.068
WP	0.070	0.121	0.079	0.132	0.147	0.221	0.132	0.157	0.304	0.100	0.067	0.089

Table 2.3 Genetic distances between 37 breeds estimated from the B blood group system ("independent loci" method below the diagonal, "reduced haplotype" method above the diagonal) – continued.

	GB	GL	GU	HF	PH	HL	IC	JS	KY	LM	LO	LR	MA
AA	0.071	0.257	0.142	0.040	0.043	0.034	0.029	0.189	0.018	0.133	0.011	0.013	0.128
AY	0.137	0.339	0.220	0.007	0.009	0.090	0.083	0.270	0.004	0.211	0.010	0.011	0.206
BB	0.092	0.289	0.177	0.023	0.025	0.064	0.046	0.233	0.007	0.174	0.007	0.006	0.165
BG	0.035	0.184	0.076	0.105	0.109	0.001	0.019	0.106	0.075	0.058	0.061	0.058	0.070
BA	0.031	0.059	0.005	0.236	0.240	0.043	0.060	0.019	0.199	0.004	0.180	0.179	0.018
BW	0.105	0.306	0.189	0.018	0.020	0.065	0.058	0.239	0.007	0.180	0.007	0.005	0.175
BS	0.129	0.046	0.044	0.389	0.394	0.165	0.184	0.020	0.355	0.050	0.333	0.337	0.063
CH	0.031	0.080	0.018	0.219	0.223	0.045	0.044	0.029	0.182	0.017	0.158	0.165	0.001
CI	0.229	0.102	0.116	0.501	0.505	0.261	0.293	0.069	0.467	0.115	0.447	0.448	0.146
DX	0.040	0.087	0.016	0.231	0.236	0.034	0.063	0.018	0.195	0.001	0.175	0.174	0.023
HO	0.100	0.297	0.183	0.019	0.021	0.069	0.051	0.239	0.004	0.181	0.004	0.006	0.170
GA	0.073	0.263	0.147	0.040	0.043	0.036	0.034	0.193	0.020	0.136	0.015	0.014	0.136
GB		0.091	0.030	0.164	0.168	0.034	0.014	0.081	0.129	0.044	0.112	0.111	0.034
GL	0.123		0.034	0.367	0.371	0.169	0.154	0.070	0.330	0.077	0.311	0.312	0.067
GU	0.086	0.078		0.251	0.255	0.064	0.066	0.022	0.215	0.013	0.196	0.197	0.016
HF	0.193	0.233	0.189		0.000	0.124	0.106	0.304	0.007	0.247	0.016	0.023	0.235
PH	0.196	0.240	0.187	0.005		0.129	0.110	0.308	0.008	0.251	0.019	0.026	0.240
HL	0.117	0.182	0.159	0.066	0.079		0.023	0.088	0.092	0.044	0.077	0.075	0.059
IC	0.050	0.183	0.077	0.215	0.216	0.155		0.116	0.073	0.071	0.057	0.061	0.052
JS	0.210	0.095	0.118	0.270	0.278	0.223	0.261		0.269	0.012	0.247	0.252	0.029
KY	0.153	0.137	0.128	0.172	0.190	0.164	0.156	0.201		0.211	0.003	0.009	0.198
LM	0.102	0.131	0.035	0.181	0.183	0.154	0.104	0.118	0.134		0.190	0.190	0.021
LO	0.219	0.236	0.191	0.164	0.175	0.157	0.239	0.264	0.165	0.216		0.009	0.174
LR	0.226	0.178	0.181	0.291	0.307	0.261	0.246	0.206	0.092	0.164	0.250		0.181
MA	0.100	0.096	0.077	0.168	0.177	0.136	0.115	0.143	0.055	0.080	0.189	0.103	
MC	0.208	0.167	0.151	0.264	0.277	0.233	0.220	0.188	0.142	0.125	0.312	0.140	0.083
MR	0.080	0.139	0.091	0.104	0.112	0.073	0.081	0.195	0.083	0.090	0.187	0.184	0.047
MU	0.053	0.142	0.075	0.123	0.126	0.082	0.046	0.224	0.114	0.099	0.145	0.202	0.093
PM	0.049	0.094	0.053	0.148	0.151	0.089	0.076	0.125	0.126	0.062	0.149	0.178	0.075
RM	0.143	0.154	0.087	0.265	0.270	0.229	0.164	0.138	0.155	0.050	0.245	0.138	0.103
SA	0.092	0.148	0.037	0.148	0.145	0.127	0.063	0.187	0.165	0.053	0.213	0.246	0.109
SH	0.108	0.126	0.103	0.144	0.154	0.112	0.131	0.160	0.074	0.083	0.167	0.119	0.065
SM	0.033	0.136	0.061	0.140	0.140	0.082	0.042	0.191	0.149	0.066	0.203	0.223	0.074
SD	0.059	0.135	0.068	0.151	0.156	0.096	0.074	0.162	0.104	0.066	0.122	0.163	0.066
SX	0.121	0.202	0.111	0.122	0.123	0.128	0.101	0.252	0.136	0.128	0.119	0.228	0.129
WB	0.056	0.097	0.075	0.130	0.138	0.078	0.083	0.169	0.075	0.091	0.140	0.135	0.030
WP	0.177	0.201	0.155	0.141	0.155	0.114	0.187	0.230	0.121	0.176	0.022	0.194	0.135

Table 2.3 Genetic distances between 37 breeds estimated from the B blood group system ("independent loci" method below the diagonal, "reduced haplotype" method above the diagonal) – continued.

	MC	MR	MU	PM	RM	SA	SH	SM	SD	SX	WB	WP
AA	0.390	0.009	0.003	0.144	0.299	0.157	0.016	0.030	0.061	0.001	0.045	0.029
AY	0.471	0.014	0.032	0.223	0.381	0.236	0.004	0.086	0.128	0.015	0.107	0.085
BB	0.435	0.001	0.014	0.185	0.343	0.191	0.005	0.052	0.093	0.008	0.068	0.050
BG	0.291	0.051	0.012	0.069	0.203	0.088	0.070	0.009	0.014	0.027	0.020	0.010
BA	0.131	0.163	0.100	0.003	0.067	0.007	0.193	0.042	0.015	0.135	0.036	0.049
BW	0.441	0.007	0.016	0.192	0.350	0.205	0.004	0.061	0.100	0.006	0.080	0.059
BS	0.027	0.321	0.248	0.038	0.005	0.044	0.349	0.166	0.115	0.288	0.147	0.175
CH	0.164	0.152	0.095	0.015	0.093	0.027	0.176	0.042	0.023	0.122	0.026	0.043
CI	0.004	0.433	0.356	0.101	0.013	0.111	0.462	0.268	0.205	0.398	0.250	0.279
DX	0.138	0.161	0.094	0.004	0.072	0.021	0.189	0.043	0.013	0.127	0.040	0.048
HO	0.441	0.001	0.017	0.191	0.350	0.198	0.003	0.057	0.099	0.008	0.073	0.055
GA	0.394	0.011	0.004	0.149	0.302	0.163	0.017	0.035	0.065	0.002	0.052	0.033
GB	0.238	0.095	0.053	0.047	0.156	0.038	0.122	0.014	0.017	0.082	0.009	0.016
GL	0.114	0.290	0.229	0.064	0.073	0.029	0.324	0.143	0.112	0.272	0.121	0.152
GU	0.122	0.178	0.118	0.009	0.062	0.002	0.209	0.053	0.028	0.155	0.041	0.061
HF	0.503	0.026	0.057	0.258	0.415	0.268	0.008	0.115	0.162	0.035	0.134	0.114
PH	0.507	0.029	0.060	0.263	0.419	0.272	0.010	0.119	0.166	0.038	0.138	0.118
HL	0.266	0.066	0.021	0.054	0.180	0.074	0.087	0.009	0.009	0.039	0.019	0.011
IC	0.300	0.049	0.022	0.077	0.211	0.078	0.068	0.007	0.023	0.038	0.004	0.005
JS	0.070	0.237	0.164	0.007	0.024	0.027	0.263	0.097	0.055	0.199	0.086	0.106
KY	0.471	0.009	0.032	0.222	0.380	0.231	0.001	0.082	0.127	0.016	0.099	0.080
LM	0.120	0.177	0.109	0.001	0.058	0.018	0.205	0.052	0.019	0.143	0.047	0.058
LO	0.451	0.010	0.025	0.201	0.359	0.213	0.003	0.068	0.110	0.010	0.081	0.064
LR	0.454	0.007	0.024	0.202	0.362	0.212	0.007	0.066	0.108	0.011	0.084	0.061
MA	0.154	0.168	0.111	0.017	0.086	0.025	0.193	0.052	0.031	0.139	0.033	0.054
MC		0.437	0.361	0.107	0.014	0.119	0.465	0.275	0.211	0.402	0.258	0.288
MR	0.151		0.015	0.188	0.346	0.192	0.007	0.052	0.094	0.010	0.069	0.051
MU	0.193	0.069		0.120	0.270	0.131	0.029	0.018	0.043	0.006	0.033	0.019
PM	0.137	0.066	0.065		0.048	0.013	0.216	0.059	0.025	0.155	0.050	0.065
RM	0.097	0.154	0.164	0.084		0.062	0.374	0.189	0.132	0.310	0.173	0.200
SA	0.189	0.081	0.072	0.066	0.139		0.225	0.062	0.035	0.170	0.050	0.070
SH	0.141	0.069	0.076	0.091	0.118	0.135		0.077	0.121	0.014	0.094	0.075
SM	0.176	0.039	0.041	0.041	0.138	0.043	0.095		0.009	0.038	0.004	0.002
SD	0.159	0.051	0.065	0.025	0.090	0.074	0.078	0.044		0.070	0.011	0.013
SX	0.225	0.108	0.039	0.094	0.183	0.119	0.115	0.093	0.093		0.055	0.037
WB	0.142	0.034	0.047	0.050	0.135	0.087	0.061	0.040	0.035	0.100		0.005
WP	0.259	0.129	0.111	0.118	0.211	0.170	0.135	0.152	0.084	0.101	0.092	

be inflated by the fact that the British breeds appear to be fixed for one 'allele', the null allele.

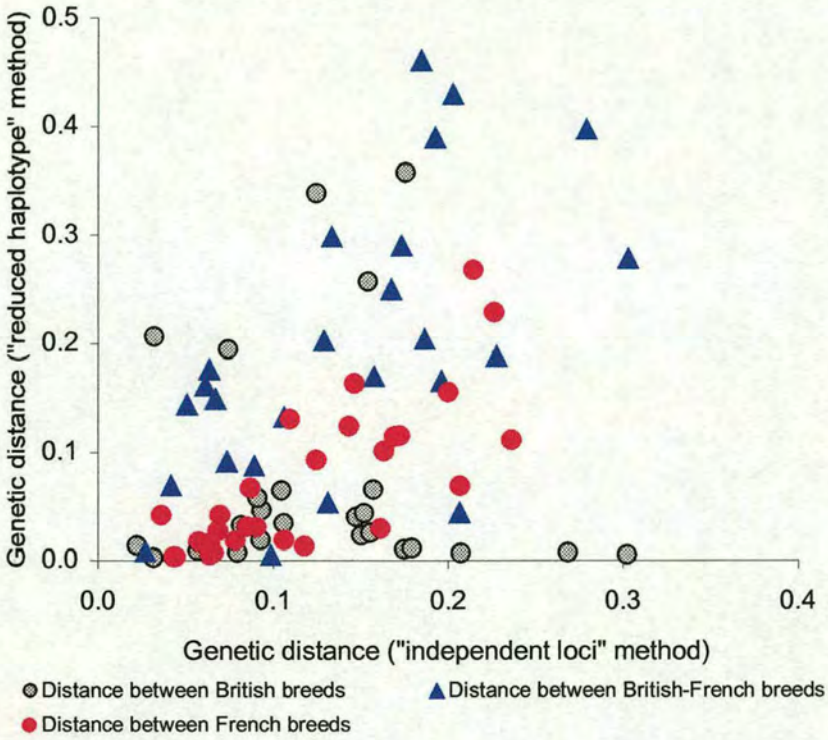

Figure 2.2 Relationship between genetic distances (among 37 European cattle breeds) estimated from the B blood type system by two different methods: 1) treating antigenic factors as independent loci or 2) using the "reduced haplotype" method ofGrosclaude et al (1990). Only a subset of the distances are shown, with points coded according to whether the distance is between two British breeds, two French breeds or a British and a French breed.

Figures 2.3 ("independent loci" method) and 2.4 ("reduced haplotype" method) show the aligned multidimensional scaling plots. Although the relative positions of the breeds are similar, the "reduced haplotype" method results in a more condensed grouping. This is presumably the result of loss of information when animals are considered to have a null phenotype, which is interpreted in the genetic distance calculation as the breeds being similar.

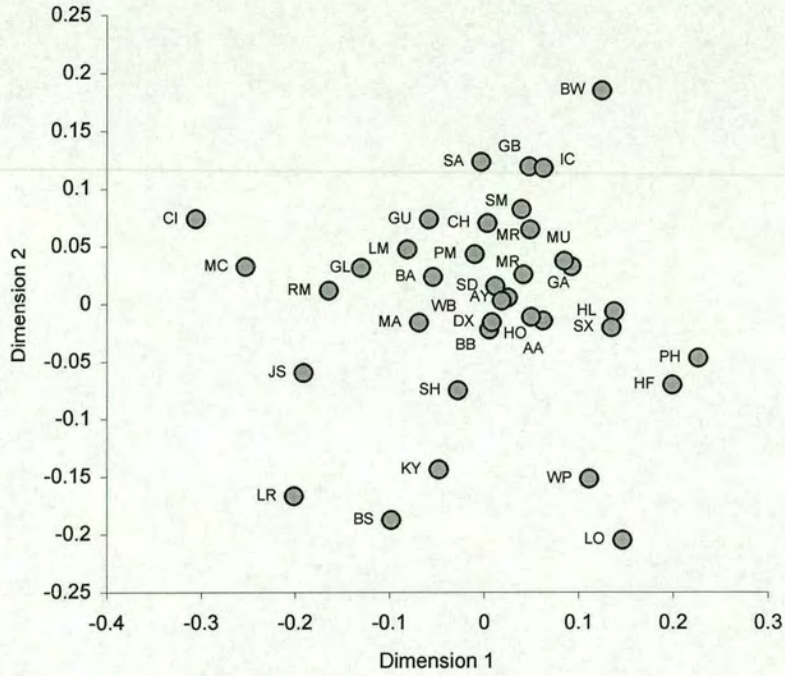

Figure 2.3 Genetic distances between breeds represented using multidimensional scaling. Distances were estimated from the B system, defined as 18 independent loci ("independent loci" method).

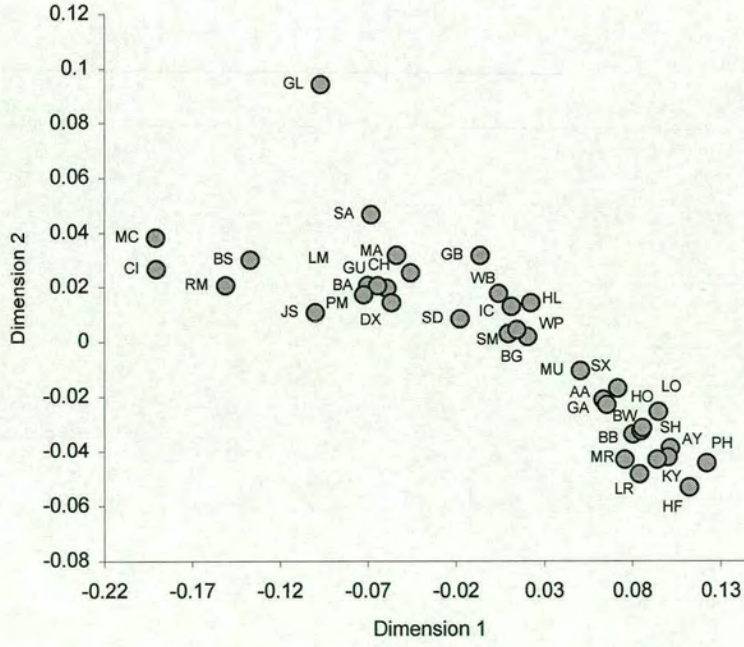

Figure 2.4 Genetic distances between breeds represented using multidimensional scaling. Distances were estimated from the B system, defined as a single locus with six alleles ("reduced haplotype" method).

2.2.3 Conclusions

The "independent loci" and "reduced haplotype" methods result in genetic distance estimates that are significantly positively correlated. The "independent loci" method, which treats linked loci as though they were independent, may have the advantage that information is preserved, and no spurious similarities between breeds are generated by the grouping of genotypes. Ideally, however, the covariances between the gene frequencies at linked loci should be accounted for in the estimate of genetic distance (e.g. the covariance matrix can be incorporated into the calculation of distance between the populations see Manly 1986 p63). The next section describes the estimation of the linkage disequilibrium or covariances between linked loci of the B red cell antigen system in nine breeds, and compares the patterns of linkage disequilibrium in the different breeds.

2.3 Gene frequency correlations between linked factors of the cattle B blood type system

2.3.1 Methods

Data from nine breeds, for which more than 300 animals per breed had been sampled, were selected; Blonde d'Aquitaine (304 animals), Belgian Blue (1242), Charolais (1569), Hereford (1034) and Poll Hereford (1079), Holstein-Friesian (7778), Jersey (460), Limousin (1813) and Simmental (928). In this analysis the different subtypes of antigenic factors were distinguished from each other, and a total of 21 factors were identified that had been consistently typed over time by the Roslin Cattle Blood Typing Service. These factors were B_1, G_1, I_1, I_2, K_1, O_1, P_1, Q_2, T_1, Y_2, A'_1, D', F'_1, G'_1, I'_1, J'_1, K', O'_1, P'_1, Q', and Y'. Linkage disequilibria and squared correlations between all pairs of these factors were calculated for each of the nine breeds separately.

Each antigenic factor was assumed to represent a locus with two alleles, one coding for the presence (A) of the antigen and one for its absence (a), with 'presence' being dominant to 'absence' (Stormont 1955; Neimann-Sorensen 1956). Hardy-Weinberg

equilibrium was assumed within individual loci. For each pair of loci (A and B) the frequency *f* of each chromosome type (AB, Ab, aB, and ab) was estimated using an iterative procedure (Langley et al 1974; Hill 1974). The covariance or linkage disequilibrium (D) between loci was calculated as:

$$D = f(AB)f(ab) - f(Ab)f(aB) \qquad [2.3]$$

Squared correlations between loci were calculated as:

$$r^2 = \left(\frac{D}{\sqrt{f(A)f(B)f(a)f(b)}} \right)^2 \qquad [2.4]$$

The matrices of squared correlations for each different breed were compared by calculating the Pearson correlation coefficient between them (equation 2.2 above), and the significance of the correlation was tested using Mantel's test.

2.3.2 Results

Comparison of gene correlations among different breeds

Squared correlations between the 21 B system factors in each of the nine breeds are shown in Appendix III. In most breeds the values of the squared correlations are low, being below 0.432. Higher squared correlations are seen between factors B_1G_1, $B_1O'_1$, and $G_1O'_1$ in the Hereford. Table 2.4 shows the Pearson correlation coefficients, calculated between the matrices of squared correlations between B system factors, for each pair of breeds. The value of this correlation coefficient reflects the degree of similarity between breeds, in their patterns of linkage disequilibrium between B system factors.

Table 2.4 Pearson correlation coefficients between breeds calculated from the matrices of squared correlations between 21 B system factors

	BA	BB	CH	HF	HO	JS	LM	PH
BB	0.169							
CH	0.217	0.312						
HF	0.039	0.012	-0.014					
HO	0.324	0.131	0.384	-0.051				
JS	0.266	0.252	0.097	0.065	0.075			
LM	0.434*	0.294	0.256	0.041	0.048	0.117		
PH	0.034	0.040	0.000	0.763*	0.014	0.038	0.041	
SM	0.100	0.026	0.032	-0.020	0.054	0.083	0.040	-0.020

* $p < 0.05$ (threshold set using a sequential Bonferroni test)

After the significance thresholds had been adjusted to allow for multiple testing, using a sequential Bonferroni test (Rice 1989), only the correlations between Hereford - Poll Hereford and Blonde d'Aquitaine - Limousin were found to be significantly different from zero ($p < 0.05$).

The most similar pattern of linkage disequilibrium among B system antigenic factors was observed in the Hereford and Poll Hereford breeds, the correlation between the two breeds was 0.76. In general, the linkage disequilibria between factors reflect the breed specific haplotypes that are observed, for example, correlations were found between the factors Y_2, D' and Y_2, I'_1 in the Hereford, and a frequent haplotype in this breed is $Y_2D'I'_1$.

2.3.3. Conclusions

Patterns of linkage disequilibrium between B system factors differ across breeds, which is reflected by the low (not significantly different from zero) correlations when the gene correlations or linkage disequilibria in different breeds are compared. These results confirm the observation that associations between B system factors tend to be breed specific (Rendel 1967).

Gene mapping methods have been proposed that are based on estimates of the linkage disequilibrium between loci (reviewed in Baret and Hill 1997). Linkage

disequilibrium mapping relies on the fact that the disequilibrium between closely linked loci decays more slowly than that between unlinked loci. However, Litt and Jorde (1986) and Jorde et al (1994) found that the order of closely linked loci cannot necessarily be accurately predicted from the linkage disequilibrium between the loci. The results presented here also suggest that it would be difficult to predict the order of loci in the cattle red cell antigen B system from the linkage disequilibrium between the loci, as the predicted order would vary from breed to breed. Linkage disequilibrium between physically linked loci is expected to decay over time in proportion to the recombination fraction between loci. Loci which are loosely linked will therefore reach equilibrium more quickly than loci which are tightly linked. However, the linkage disequilibrium will also be influenced by the population's history, e.g. changes in effective population size, and the effects of migration, selection and random drift. The resulting variation in disequilibrium in different breeds explains the existence of breed specific haplotypes. If the order of genes and recombination rates between loci were known unequivocally from linkage studies, then the linkage disequilibrium between them could potentially provide valuable insights into the histories of the different breeds.

2.4 Discussion

Two methods of treating the linked loci of the B blood type system, when estimating genetic distances between breeds, were compared. The "independent loci" method, which as the name suggests treated the B system as eighteen independent diallelic loci, was more informative than the "reduced haplotype" method, which defined the B system as a single locus with six alleles. A drawback of the "reduced haplotype" method was that a large number of phenotypes were pooled into a "null" category, and this generated a spurious similarity between breeds. The criticism of the "independent loci" method might be that there is redundancy of information, with two correlated loci both perhaps contributing the same amount of information to the genetic distances as a third locus that is independent of all the others. The most accurate way to approach the problem of using linked loci for genetic distance calculations, would be to incorporate

the covariances between loci into the measure of genetic distance. Such a measure can only be calculated if the covariance matrix is the same for all populations (Manly 1986), and this is estimated by calculating a pooled covariance matrix across all populations. In practice, it has been shown that distance measures utilising the covariances between variables are superior to measures that do not, only when the covariances are accurately known (Manly 1986).

In the particular case of the B blood type system the covariances between loci were generally small, and the larger covariances were not consistently large across all breeds. The "independent loci" method was also more highly correlated with the method that identified all haplotypes ("full haplotype" method) than with the "reduced haplotype" method. In these circumstances it seems acceptable to use a distance measure which does not make use of the covariances between the loci, such as that of Reynolds et al (1983). It also seems acceptable to utilise information from all the loci within each cattle blood type system, since the linkage disequilibrium between loci is small, and treating the loci as though they are independent gives similar estimates of genetic distance to those obtained from haplotypes. Ignoring linkage between loci may, however, have a greater impact on estimates of the variance than on the observed value of genetic distance, and this is investigated in the following chapter (chapter 3).

When the precise order and recombination rates between loci are known, then the use of linked genes for population studies can provide valuable insights into population history, particularly for dating divergence times (Tishkoff et al 1996). Differences in the disequilibria between loci in different populations could also provide information on whether populations have been through bottlenecks or admixture between populations has occurred. The use of linked loci for studies of population structure and genetic diversity is not, therefore, necessarily disadvantageous. However, the best markers to use for this type of study would be those for which the genetic map is well defined, i.e. the precise location and distances between loci are known. The existence of genome maps for a large number of domesticated species, both plant and animal, provides new opportunities for the assessment of genetic diversity, allowing

markers to be selected according to their map position. For example, new methods of measuring diversity, utilising information on the map position of markers and relating genetic distances to morphological traits, are currently being developed for measuring the distinctiveness of maize varieties (Dillmann et al 1997). This type of approach to the measurement of genetic diversity will allow genetic information to be related to information on phenotypic differences between breeds or strains.

Chapter 3

Estimating the sampling variance of genetic distance

3.1 Introduction

One of the key aims in studies of livestock breed diversity is the identification of genetically unique populations, and "taxonomic distinctness" has been suggested as a criterion by which these populations can be identified (Hall and Bradley 1995). Genetic distance is frequently used to estimate the evolutionary difference between breeds, and for the construction of phylogenetic trees of breed relationships (Grosclaude et al 1990; MacHugh et al 1994). The sampling variance of genetic distance gives an indication of the precision of the distance estimate, and the magnitude of the variance affects the accuracy of phylogenetic tree construction and the chances of recovering true breed relationships (Nei et al 1983; Goldstein and Pollock 1994; Tajima and Takezaki 1994).

If genetic distance is thought of as a measure of overall genomic difference between two populations then two sampling processes are involved in its estimation; sampling of loci from the genome and sampling of individuals (genes) from the populations (Nei 1987). The sampling variance can therefore be partitioned into two components, the interlocus and intralocus variance. The intralocus variance is a function of the number of individuals sampled from the populations, and the allele frequencies at the loci sampled. Interlocus variance arises because of differences between loci caused by drift, and variation in the mutation rate. Slatkin and Arter (1991) defined the sources of variation as; sampling variation (a result of the sampling process when the data are collected), stochastic variation (the result of random processes governing allele frequencies at that locus, otherwise known as drift) and parametric variation (resulting from differences in mutation rate among loci).

Weir (1996) defines the components of variance in terms of 'fixed' and 'random' populations, or statistical (sampling of genes from the population) and genetic sampling (sampling of genes between generations). He states that if the populations can be thought of as representing the whole species or breed then the 'fixed' approach can be used, however if the populations represent a random sample from a set of possible replicate populations (that have arisen from the same founder population) then they must be regarded as 'random'. In defining populations as 'random' and studying more than one locus (each locus has an independent history and to a certain extent each will represent a replicate population (Weir 1996; Falconer 1989)) an estimate of the drift variance can be obtained and inferences made about the evolution of the populations. If statements are to be made about some future sample, then the variation between replicate populations must be taken into account (Weir 1996).

3.1.1 Nei's formula for the sampling variance of genetic distance

Nei (1987) derived a formula for the variance of his standard genetic distance, under the assumption of linkage equilibrium among all loci, and Nei and Roychoudhury (1974) obtained the asymptotic variance when sample size is large. The formula shows how the components of the variance can be partitioned into the interlocus variance, which is due to genetic sampling between generations (Weir 1996), and the intralocus variance, which is due to statistical sampling of individuals from the populations (Weir 1996). Nei's standard genetic distance (D) is given by:

$$D = -\ln\left(\frac{J_{xy}}{\sqrt{J_x J_y}}\right) \qquad [3.1]$$

where $J_x = \sum_{k=1}^{r} \frac{j_x}{r}$, $J_y = \sum_{k=1}^{r} \frac{j_y}{r}$, $J_{xy} = \sum_{k=1}^{r} \frac{j_{xy}}{r}$,

and $j_x = \sum x_i^2$, $j_y = \sum y_i^2$, $j_{xy} = \sum x_i y_i$,

and where r = number of loci, x_i = frequency of allele i in population 1, y_i = frequency of allele i in population 2, and $\frac{J_{xy}}{\sqrt{J_x J_y}}$ is the gene identity (probability that two alleles picked at random, one from each population, are identical by descent).

The sampling variance is composed of two components (interlocus and intralocus variance):

$$V(\hat{D}) = V_r(\hat{D}) + V_s(\hat{D})$$

The overall variance is given by:

$$V(\hat{D}) = \frac{V(\hat{J}_x)}{4\hat{J}_x^2} + \frac{V(\hat{J}_y)}{4\hat{J}_y^2} + \frac{V(\hat{J}_{xy})}{\hat{J}_{xy}^2} + \frac{Cov(\hat{J}_x, \hat{J}_y)}{2\hat{J}_x\hat{J}_y} - \frac{Cov(\hat{J}_x, \hat{J}_{xy})}{\hat{J}_x\hat{J}_{xy}} - \frac{Cov(\hat{J}_y, \hat{J}_{xy})}{\hat{J}_y\hat{J}_{xy}}$$

[3.2]

$V(\hat{J}_x)$ and $V(\hat{J}_y)$ are the variance of average heterozygosity in each population

$$V(J_x) = \sum_{k=1}^{r} \left(j_{x(k)} - J_x\right)^2 \Big/ (r-1)$$

and $Cov(\hat{J}_x, \hat{J}_y)$, etc can be calculated, since $\hat{J}_x$, $\hat{J}_y$ and $\hat{J}_{xy}$ are the average of single locus gene identities. For example,

$$Cov(\hat{J}_x, \hat{J}_y) = \sum_{k=1}^{r} (\hat{j}_{xk} - \hat{J}_x)(\hat{j}_{yk} - \hat{J}_y) / \{r(r-1)\}$$

The intralocus component of variance ($V_s(\hat{D})$) is given by:

$$V_s(\hat{D}) = \left[\frac{\sum V_s(\hat{j}_x)}{4\hat{J}_x^2} + \frac{\sum V_s(\hat{j}_y)}{4\hat{J}_y^2} + \frac{\sum V_s(\hat{j}_{xy})}{4\hat{J}_{xy}^2} - \frac{\sum Cov_s(\hat{j}_x, \hat{j}_{xy})}{\hat{J}_x\hat{J}_{xy}} - \frac{\sum Cov_s(\hat{j}_y, \hat{j}_{xy})}{\hat{J}_y\hat{J}_{xy}} \right] \Big/ r^2$$

[3.3]

where for each locus

$$V_s(\hat{j}_x) = \frac{2(n-1)}{n^3}\left[(3-2n)j_x^2 + 2(n-2)\sum x_i^3 + j_x\right]$$

$$V_s(\hat{j}_y) = \frac{2(n-1)}{n^3}\left[(3-2n)j_y^2 + 2(n-2)\sum y_i^3 + j_y\right]$$

$$V_s(j_{xy}) = \left[(1-n_x-n_y)\left(\sum x_i y_i\right)^2 + (n_x-1)\sum x_i^2 y_i + (n_y-1)\sum x_i y_i^2 + \sum x_i y_i\right] \Big/ (n_x n_y)$$

$$Covs(j_x, j_{xy}) = 2(n_x-1)\left[\sum x_i^2 y_i - \left(\sum x_i^2\right)\left(\sum x_i y_i\right)\right] \Big/ n_x^2$$

$$Covs(j_y, j_{xy}) = 2(n_y-1)\left[\sum x_i y_i^2 - \left(\sum y_i^2\right)\left(\sum x_i y_i\right)\right] \Big/ n_y^2$$

Formulae [3.2] and [3.3] show that the overall sampling variance depends on the variance of average heterozygosity within each population, and the intralocus component of variance depends on the number of genes sampled ($2n$ where n is the number of individuals sampled) and the gene frequencies (x and y). Equation [3.3] also implies that the intralocus variance is influenced by the number of loci sampled. Chakraborty (1985) has shown that formula [3.2] gives quite an accurate value for the overall sampling variance even if the number of loci sampled is as few as twenty.

3.1.2 Bootstrapping for estimating the sampling variance of genetic distance

Genetic distance is a complex function, it does not have simple statistical properties and derivation of the variance is not a straightforward procedure. Nei's standard genetic distance is the only distance measure for which a formula has been derived. Numerical sampling methods, such as jackknifing and bootstrapping, have been suggested as a means of estimating the variance for other measures of genetic distance (Meuller 1979; Sanchez et al 1995).

The bootstrap was introduced as a computer-based method for estimating standard errors (Efron and Tibshirani 1993). In the context of genetic studies, bootstrapping

has been suggested as a method for estimating confidence intervals around F-statistics and gene frequencies (Weir 1996). Its use has also been demonstrated for estimating confidence intervals in linkage analysis (Chiano and Yates 1994) and QTL mapping (Visscher, Thompson and Haley 1996). Felsenstein (1985) introduced bootstrapping as a method for assessing confidence in the structure of phylogenies. His method resamples characters or loci, a new tree is constructed from the resampled characters and the results are then combined in a consensus tree. A p value is calculated, this value being the number of times a particular clade (group of populations) is represented among the replicate trees, which is interpreted as a measure of repeatability ("probability that a specified group will be found in an analysis of an independent sample of characters", (Hillis and Bull 1993)). When genetic markers are the characters in question, then the method investigates the effect on the phylogeny of sampling different loci from the genome, effectively the interlocus variance is explored. When distance matrix methods of phylogeny inference are employed, the genetic distances are bootstrapped and a new tree generated from the distance matrix of each bootstrap replicate. The intralocus variance is not taken into account by Felsenstein's method (in his PHYLIP program (Felsenstein 1995), loci are resampled from the array of allele frequencies), and it is assumed that the allele frequencies have been estimated without error.

Bootstrapping works by drawing random samples of the same size as the original sample from that sample. If the original sample consists of n observations a bootstrap sample is a set of n observations drawn at random, with replacement, from this set so that every one of the original observations has an equal chance of being chosen. Some of the original sample elements will not, therefore, appear in any particular bootstrap sample while some may appear many times. For each bootstrap data set x^* a bootstrap replicate of the parameter of interest θ is calculated (in this instance θ is the genetic distances between breeds).

The bootstrap estimate of the standard error of the statistic is obtained by calculating the standard deviation of the replicates, given by:

$$\hat{s}e_B = \left\{ \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \hat{\theta}^*(.) \right]^2 / (B-1) \right\}^{1/2}$$

where $\hat{\theta}^*(.) = \sum_{b=1}^{B} \hat{\theta}^*(b) / B$

$\hat{\theta}^*(b)$ is the value of θ in the *bth* bootstrap replicate

and B = number of bootstrap replicates

(from Efron and Tibshirani 1986)

One of the potential advantages of the bootstrap is that the sampling strategy can be adjusted to account for correlated data (Efron and Tibshirani 1986; Liu and Singh 1992), for example, relationships between linked genes could be taken into account by sampling the loci together. This makes bootstrapping an appealing method for application to the cattle blood type data, because of the linkage between loci that is involved.

The two components of the sampling variance of genetic distance can be estimated by the use of two different bootstrap sampling strategies. By resampling individual genotypes the intralocus variance can be estimated, and by resampling loci the interlocus variance is estimated. However, it has been suggested that the generally small number of loci assayed in any one study means that bootstrapping by resampling loci may not be an effective strategy for estimating the interlocus variance (Brown 1994; Van Dongen 1995). When a small number of loci is sampled a discrete distribution is obtained, with the genetic distance only able to take a limited number of values (Van Dongen 1995). Since the information used to approximate the bootstrap distribution comes from the observed data, it follows that if the number of observations is very small then the data may not contain sufficient information about the underlying distribution. A further assumption of the bootstrap is that the random variable being sampled is independently and identically distributed. Van Dongen (1995) suggests that if loci are sampled this requirement is also not met, since different loci are subject to different mutation rates and, therefore,

cannot be considered to be drawn from the same distribution. Brown (1994) also notes that the assumption of independence will not be valid if the markers are linked.

Estimates of the standard error of genetic distance are necessary, as an indication of the accuracy of the distance measurement and the confidence with which genetic relationships can be inferred. However, a theoretical estimate of the standard error has only been derived for Nei's standard genetic distance. There are many other measures of genetic distance (see chapter 1), and for the particular problem under study it may be desired to use a measure other than Nei's standard distance. Additionally, the theoretical estimate of standard error for Nei's genetic distance is based on the assumption that the loci are independent (in linkage equilibrium). Bootstrapping provides an alternative way of estimating the standard error of genetic distance. It has the potential advantages of being easy to compute, and of not being limited to use under particular assumptions. However, it is necessary to establish whether the bootstrap can provide similar answers to those obtained with Nei's theoretical estimate, and to investigate how the inclusion of linked loci affects the estimate of standard error.

The objective of this chapter was to explore the use of bootstrapping as a method for estimating the sampling variance of genetic distance. The intralocus and interlocus variances of genetic distance were estimated by bootstrapping and compared with values obtained from Nei's forumula, using simulated data. The effect of accounting for linkage among loci by sampling linked loci as a group was investigated using blood type data from seven cattle breeds.

3.2 Methods

3.2.1 Simulated data

Allele frequencies at 50 independent diallelic loci in two populations were randomly generated, by drawing random numbers from a uniform distribution between zero

and one. The first allele at each locus was assigned the random number (p) as its frequency, the second allele at the locus had a frequency of 1-p. Table 3.1 shows the allele frequencies assigned to each population for the first ten loci.

Table 3.1 Allele frequencies generated at the first ten diallelic loci in two simulated populations

		Locus									
Population	Frequency	1	2	3	4	5	6	7	8	9	10
1	Allele 1	0.80	0.23	0.04	0.15	0.60	0.81	0.02	0.62	0.97	0.52
	Allele 2	0.20	0.77	0.96	0.85	0.40	0.19	0.98	0.38	0.03	0.48
2	Allele 1	0.79	0.49	0.89	0.24	0.04	0.65	0.23	0.33	0.09	0.71
	Allele 2	0.21	0.51	0.11	0.76	0.96	0.35	0.77	0.67	0.91	0.29

The allele frequencies generated at each of the 50 loci were taken as the true frequencies. The interlocus variance for 10, 15, 20, 30, 40 and 50 loci was estimated using bootstrapping by resampling loci (see section 3.2.3 below), and the estimates were compared with those obtained using Nei's formula (see equation 3.2) when the population sample was very large, i.e. more than 100,000 individuals (computed using the program DISPAN (Ota 1993)).

In addition, five data sets of 15, 25, 50, 100 or 500 individual genotypes were simulated from the true allele frequencies for each of the first 10 and first 20 loci. Allele frequencies were estimated from each data set and these frequencies were substituted into Nei's formula (computed using the DISPAN program (Ota 1993)) to obtain estimates of Nei's (1972) standard genetic distance between the two populations and its sampling variance. Two different bootstrap sampling strategies (see section 3.2.3 below) were also used on each data set, resampling loci (figure 3.2) and resampling individual genotypes (figure 3.1), in order to estimate the two different components (interlocus and intralocus) of the sampling variance of Nei's (1972) genetic distance. For each data set 200 bootstrap replicates were generated,

and the variance components estimated as the variance of the genetic distance over the 200 replicates.

3.2.2 Real data

Real blood type data from 7 cattle breeds (Aberdeen Angus, Charolais, Hereford, Holstein-Friesian, Jersey, Kerry and Simmental) were used to investigate the effect of accounting for linkage between loci when bootstrapping to obtain an estimate of the sampling variance of genetic distance. Allele frequencies were estimated at 31 loci, which formed 9 linkage groups. The number of loci in each group varied between 1 and 18, but where loci were linked it was very tightly (0.7 cM or less between loci) – see chapter 2 for details of cattle blood groups and their genetic structure. The genetic distance between breeds was estimated using the measure of Reynolds et al (1983). This measure of genetic distance was chosen, rather than Nei's distance, because it was felt to be the most appropriate measure for the analysis of closely related populations, such as cattle breeds, and was the measure used in the analyses presented in chapters 4 and 5. The two components of variance were estimated by resampling loci (interlocus variance, figure 3.2) or resampling individual genotypes (intralocus variance, figure 3.1), each with two variations either i) linked loci were sampled together as a group or ii) loci were assumed to be independent and were sampled separately.

3.2.3 Bootstrap sampling strategies

The bootstrap estimate of the intralocus variance was computed by sampling, with replacement, the same number of individual genotypes as were originally observed. When the loci were assumed to be unlinked, sampling of loci was carried out independently, so within an individual record (an individual's genotype at all loci) each locus was sampled independently of the others. This resulted in individual records in the bootstrap sample being composed of completely new genotypes, so for example, each allele in the bootstrap record could have come from different animals

in the original sample. For linked loci the genotype at these loci was sampled as a single unit. In this case an individual genotype at a linked group of loci would be sampled from one individual in the original sample. For each bootstrap replicate, a count of the phenotypes was made, gene frequencies estimated and genetic distance between breeds calculated. The procedure is illustrated in figure 3.1. The estimate of the intralocus variance or standard error was obtained from the variance (or standard error) of the genetic distance over all bootstrap replicates.

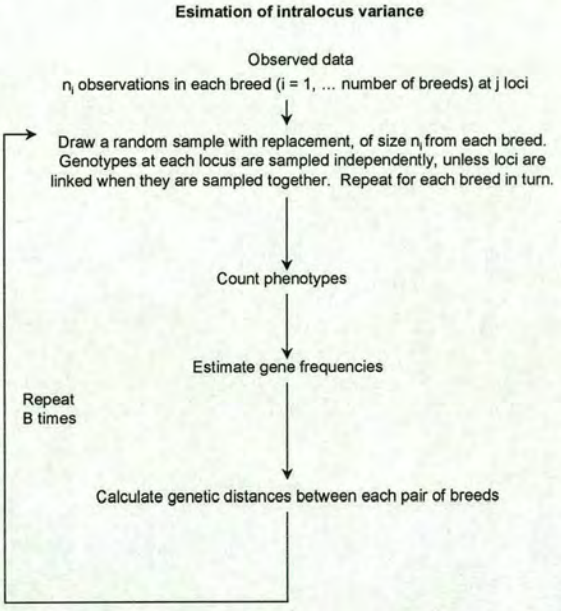

Figure 3.1 Flow diagram of the bootstrap strategy used for estimation of the intralocus component of the sampling variance of genetic distance

The bootstrap estimate of the interlocus variance was computed by resampling loci or, more precisely, the allele frequencies at the loci. Allele frequencies at each locus were estimated from the observed data. The frequencies were then randomly sampled, with replacement, for the same number of loci as were originally observed. Genetic distances were calculated from this new array of allele frequencies (see figure 3.2), and the interlocus variance of the genetic distance was obtained by calculating the variance of the genetic distance over the bootstrap replicates.

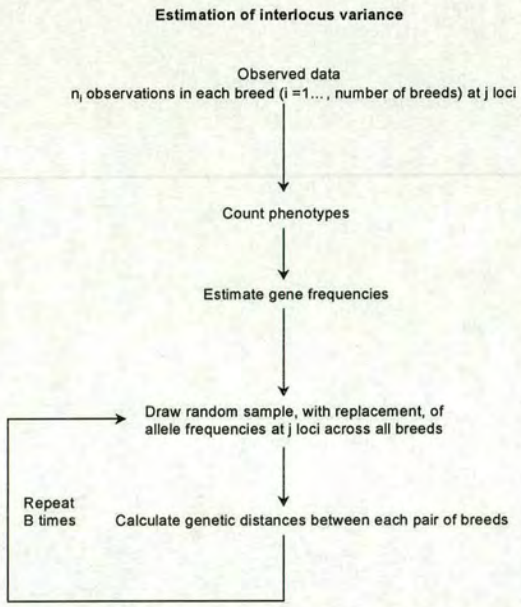

Figure 3.2 Flow diagram of the bootstrap strategy used for estimation of the interlocus component of the sampling variance of genetic distance

3.3 Results

3.3.1 Comparison of bootstrap method of estimating the standard error of genetic distance with Nei's formula

Table 3.2 and figure 3.3 compare the bootstrap estimate of standard error (sampling over loci) and Nei's formula value for different numbers of loci ranging between 50 and 5, when a large or infinite number of individual animals have been sampled. The difference between the two estimates is less than 0.005 for between 50 and 20 loci, but increases as the number of loci decreases to 15 or below, with the bootstrap value being less than the formula value. The curve's lack of smoothness, particularly noticeable when the number of loci is less than 15, is due to sampling error. Only a single replicate was used to estimate the standard errors and the use of more replicates would have given a smoother curve.

Table 3.2 Comparison between the bootstrap estimate of standard error (sampling over loci) and Nei's formula value for different numbers of loci, when a large or infinite number of individual animals have been sampled.

No. loci	50	40	30	20	15	10	5
Bootstrap s.e. (1)	0.0663	0.0784	0.0885	0.1303	0.1703	0.1640	0.2238
Formula s.e. (2)	0.0633	0.0765	0.0881	0.1279	0.1647	0.1769	0.2590
Difference (1)-(2)	0.0030	0.0019	0.0004	0.0024	0.0056	-0.0129	-0.0352

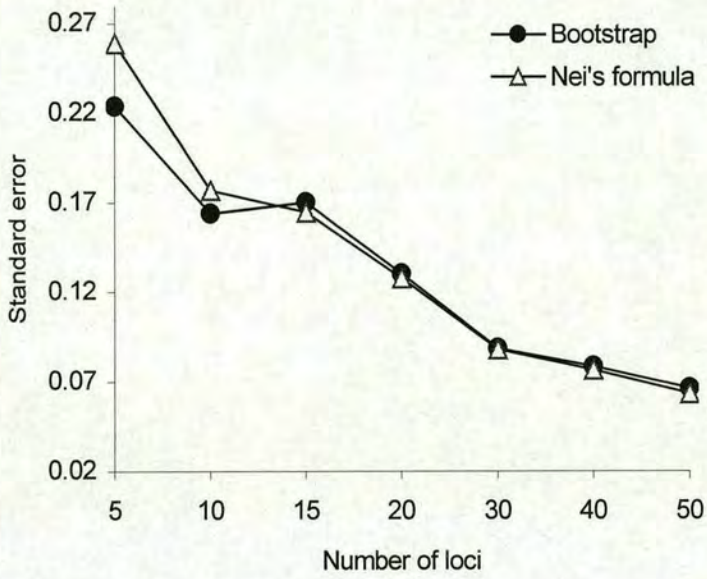

Figure 3.3 Comparison between bootstrap estimate of standard error and Nei's formula value for different numbers of loci.

Tables 3.3 and 3.4 show the effect on the standard error when different numbers of individuals are sampled from a population. Between 500 and 15 individuals were sampled at 10 loci (table 3.3) and 20 loci (table 3.4). The intralocus component of the variance, shown as intralocus standard error (s.e.) in the tables, increases as the number of individuals sampled decreases. Differences in the interlocus variance for each group of individuals are caused by sampling, i.e. the allele frequencies at the loci have been estimated from a finite or small sample of individuals from the population. For 20 loci the difference between the overall bootstrap estimate of the standard error, calculated by summing the interlocus and intralocus variance then taking the square root, and the formula value is very small (0.005 or less). However,

for 10 loci the bootstrap estimate is always smaller than the formula estimate (between 0.01 and 0.02 smaller). This is probably due to the small number of loci that have been sampled, the results in table 3.2 (and figure 3.3) indicate that there is a larger discrepancy between the two estimates of the variance when less than 15 loci are assayed.

Table 3.3 Effect on the standard error of the genetic distance of sampling different numbers of individual animals at 10 loci. Table shows the comparison between the bootstrap estimates of inter- and intra-locus variance against Nei's formula value.

	10 loci				
No individuals sampled	500	100	50	25	15
Intralocus s.e. (bootstrap)	0.0069	0.0150	0.0250	0.0303	0.0330
Interlocus s.e. (bootstrap)	0.1622	0.1607	0.1503	0.1352	0.1742
Overall s.e. (bootstrap) (1)	0.1623	0.1614	0.1523	0.1385	0.1773
Formula s.e. (2)	0.1806	0.1833	0.1674	0.1481	0.1957
Difference (1) – (2)	-0.0183	-0.0219	-0.0151	-0.0096	-0.0184

Table 3.4 Effect on the standard error of the genetic distance of sampling different numbers of individual animals at 20 loci. Table shows a comparison between the bootstrap estimates of inter- and intra-locus variance against Nei's formula value.

	20 loci				
No individuals sampled	500	100	50	25	15
Intralocus s.e. (bootstrap)	0.0059	0.0152	0.0174	0.0300	0.0379
Interlocus s.e. (bootstrap)	0.1268	0.1317	0.1243	0.1332	0.1266
Overall s.e. (bootstrap) (1)	0.1269	0.1326	0.1255	0.1365	0.1321
Formula s.e. (2)	0.1278	0.1331	0.1307	0.1372	0.1343
Difference (1) – (2)	-0.0009	-0.0005	-0.0052	-0.0007	-0.0022

3.3.2 Effect of accounting for linkage among loci when bootstrapping

Table 3.5 shows the bootstrap means and standard errors of the genetic distances among seven European cattle breeds. Parts (a) and (b) give the intralocus standard errors, estimated by sampling individual genotypes, with either linked loci sampled together (a) or with loci sampled independently (b). Parts (c) and (d) show the interlocus standard error, estimated by sampling over loci, with either linked loci sampled together (c) or loci sampled independently (d). The relationships between

the two methods of sampling (linked loci together or all loci independently) when estimating the intralocus standard error or the interlocus standard error are illustrated in figures 3.4 and 3.5 below.

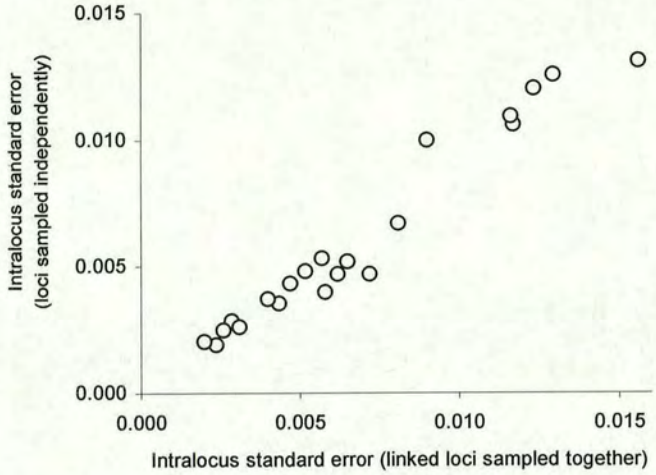

Figure 3.4 Relationship between the intralocus standard error when estimated by bootstrapping with either linked loci sampled together or loci sampled independently.

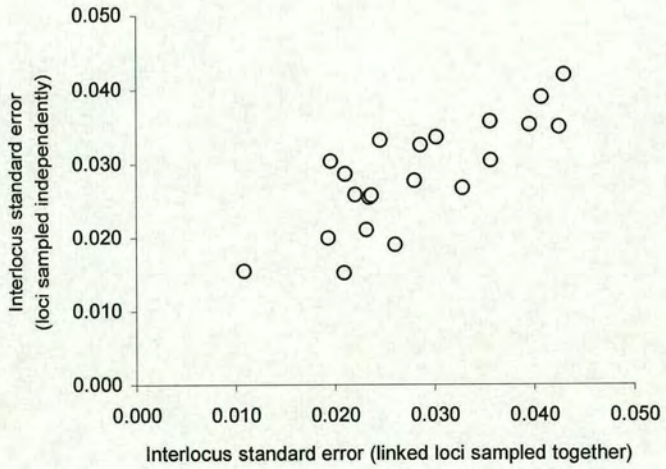

Figure 3.5 Relationship between the interlocus standard error when estimated by bootstrapping with either linked loci sampled together or loci sampled independently.

Table 3.5a Bootstrap mean genetic distance (below diagonal) and standard error (above the diagonal) when resampling individual genotypes (intralocus variance) with linked loci sampled together.

Breed	AA	CH	HO	HF	JS	KY	SM
AA		0.005	0.004	0.005	0.008	0.013	0.006
CH	0.105		0.002	0.003	0.006	0.012	0.002
HO	0.045	0.115		0.003	0.006	0.012	0.003
HF	0.129	0.136	0.130		0.007	0.012	0.004
JS	0.207	0.142	0.187	0.234		0.009	0.007
KY	0.097	0.143	0.104	0.161	0.191		0.016
SM	0.072	0.055	0.077	0.140	0.171	0.169	

Table 3.5b Bootstrap mean genetic distance (below diagonal) and standard error (above the diagonal) when resampling individual genotypes (intralocus variance) with loci sampled independently.

Breed	AA	CH	HO	HF	JS	KY	SM
AA		0.005	0.004	0.004	0.007	0.013	0.005
CH	0.105		0.002	0.003	0.004	0.011	0.002
HO	0.045	0.115		0.003	0.005	0.012	0.002
HF	0.129	0.136	0.130		0.005	0.011	0.004
JS	0.207	0.142	0.187	0.234		0.010	0.005
KY	0.098	0.143	0.105	0.160	0.191		0.013
SM	0.071	0.055	0.076	0.139	0.170	0.168	

Table 3.5c Bootstrap mean genetic distance (below diagonal) and standard error (above the diagonal) when resampling loci (interlocus variance) with linked loci sampled together.

Breed	AA	CH	HO	HF	JS	KY	SM
AA		0.019	0.021	0.030	0.040	0.033	0.011
CH	0.104		0.026	0.024	0.021	0.022	0.023
HO	0.046	0.119		0.028	0.041	0.036	0.024
HF	0.127	0.138	0.129		0.036	0.020	0.025
JS	0.198	0.132	0.188	0.224		0.043	0.043
KY	0.087	0.133	0.098	0.149	0.171		0.029
SM	0.069	0.057	0.079	0.140	0.159	0.161	

Table 3.5d Bootstrap mean genetic distance (below diagonal) and standard error (above the diagonal) when resampling loci (interlocus variance) with loci sampled independently.

Breed	AA	CH	HO	HF	JS	KY	SM
AA		0.020	0.015	0.034	0.035	0.027	0.015
CH	0.105		0.019	0.025	0.029	0.026	0.021
HO	0.044	0.115		0.028	0.039	0.030	0.026
HF	0.129	0.136	0.132		0.036	0.030	0.033
JS	0.208	0.141	0.185	0.234		0.042	0.035
KY	0.087	0.135	0.093	0.151	0.186		0.032
SM	0.071	0.054	0.076	0.139	0.168	0.158	

There is little difference between the estimates of the intralocus variance when either linked loci are sampled together or loci are sampled independently. There was more variation between the two estimates of the interlocus variance, but with no consistent pattern. Sometimes sampling linked loci together resulted in a higher estimate of the standard error than sampling loci independently and sometimes in a lower estimate. The differences were not substantial, the largest difference being 0.01 for the standard error of the distance between the Hereford and Kerry breeds.

3.4 Conclusions

Bootstrap estimates of the standard error for Nei's (1972) genetic distance were generally in good agreement with those obtained from Nei's formula. However, when the number of loci assayed is small (< 15) then the interlocus component of the variance may be underestimated by the bootstrap. There is some evidence in the literature that the bootstrap does not perform well with samples of less than 20 (Van Dongen and Backeljau 1995) and it seems likely that the empirical distribution will not accurately reflect the true distribution when based on very small samples, e.g. a sample of five loci could not be regarded as an adequate representation of the whole cattle genome, which consists of 30 chromosome pairs and hundreds of thousands of genes.

One of the potential advantages of using bootstrapping to estimate the standard error of genetic distance is that correlation or structure in the data can be accounted for. The sampling variance of genetic distance can then be estimated when, for example, loci are in linkage disequilibrium or where there is a family structure in the data. The investigation of two different sampling methods, with linked loci either sampled together or all loci sampled independently, revealed little difference in the estimates of the standard error for this particular data set. No difference was seen in the intralocus component of variance, although there were small differences in the estimates of the interlocus component of variance. If there was significant linkage disequilibrium between component loci of the cattle blood group systems then it

would be expected that treating the loci as independent would result in underestimates of the variance. The results presented in this chapter seem to support the results of the previous chapter where it was seen that linkage disequilibrium between most loci was very low, and only a few loci in each breed were actually in linkage disequilibrium. Additionally these were not the same loci in all breeds, but varied from breed to breed. The results suggest that there is not much difference in the estimates of standard error for genetic distance when bootstrapping cattle blood groups accounting for linkage, or when assuming that the component loci are independent.

Although Nei has derived a formula for the sampling variance of his standard genetic distance, no formulae have been derived for the other measures of genetic distance that exist. When using other measures of genetic distance bootstrapping provides a reasonable means of estimating the standard error (Sanchez et al 1995). If we are interested in making inferences about total genomic differences or evolutionary relationships among populations then accurate estimates of the interlocus variance are required. They are also required to compare genetic distances based on one set of loci with those based on another set of loci (Nei 1987). A sufficient number of loci (more than 15 or 20) must then be sampled for the bootstrap to provide reliable estimates of the interlocus variance. In any case, Takezaki and Nei (1996) have recommended that at least 50 loci should be sampled in order to ascertain the correct relationships, among closely related populations. However, they do not discuss whether this is an absolute value or represents a proportion of the genome. If it represents a proportion of the genome, then the number of markers required to accurately estimate relationships will vary according to the species studied (since the size of the genome varies from species to species).

Chapter 4

Genetic relationships among European cattle breeds

4.1 Introduction

The evolution of livestock breeds has been shaped by man over many generations. Geographic separation would have initially promoted divergence of populations and, in the last 100 to 200 years, many European breeds have been genetically isolated as a result of restraints imposed by pedigree herd book registration. Diverse breeds have evolved, adapted to local climates, diseases, and nutritional environments. Breeds have also been selected for different objectives, depending on the traits that were important to the local human communities. In addition to selection, genetic drift will have contributed to the differentiation of breeds and individual breeds may now possess unique combinations of genes as a result of these different evolutionary forces.

Worldwide there are more than 790 breeds of cattle, with about 270 breeds native to Europe (FAO 1995; EAAP 1993). The loss of breeds or strains would lead to a reduction in genetic diversity, which may restrict the ability of farmers to meet future agricultural requirements (Barker et al 1993; FAO 1996). However, it is not easy to predict future genetic requirements, particularly as at present very few genes controlling economically important traits in livestock have been identified, and the mapping of these genes is a still a long way from fully understanding their structure and function. Since the genetic resources required for the future are unknown, it has been suggested that one criterion that might be used to identify breeds for conservation is 'taxonomic distinctness' (Hall and Bradley 1995). The assumption underlying this criterion is that breeds having unique evolutionary histories are most likely to have special adaptations and gene combinations not found in other breeds.

By selecting for conservation those populations with unique evolutionary histories a maximum amount of diversity could be preserved (May 1990).

Analysis of biochemical and blood type polymorphisms has shown that European *Bos taurus* breeds have close genetic relationships, but are distinct from the Asian and African *Bos indicus* breeds (Baker and Manwell 1991). Previous studies to characterize relationships within the European group of cattle breeds have focused on breeds from Austria (Kidd and Pirchner 1971), Spain (Kidd et al 1980; Gonzalez et al 1987), Italy (Astolfi et al 1983) and France (Grosclaude et al 1990). Medjugorac et al (1994) examined breed relationships among eight Balkan and six other breeds, including populations of Brown Swiss, Holstein-Friesian and Jersey from the USA. Manwell and Baker (1980) defined relationships among ten major breed groups, based on a survey of protein polymorphisms published in the literature. They were not able, however, to study relationships between individual breeds and the major breed groups were defined using historical rather than genetic information.

More recently, DNA markers have been used to study cattle breed relationships; markers used include restriction fragment length polymorphisms (RFLPs) (Bradley et al 1994), randomly amplified polymorphic DNA (RAPDs) (Gwakisa et al 1994), mitochondrial DNA (Loftus et al 1994) and microsatellites (MacHugh et al 1994; Moazami-Goudarzi et al 1994; Ciampolini et al 1995; Basedow et al 1996). Microsatellites are also the type of marker chosen for a large scale international conservation project (FAO 1996), the first objective of which is to establish the extent of genetic diversity and relationships among all breeds for each of the domestic livestock species.

The accurate determination of genetic relationships among breeds may require the use of different types of genetic marker (Cunningham et al 1994). Genetic relationships established using biochemical or blood type markers will provide a useful comparison for those obtained using DNA markers. Many countries have national databases of blood typing results for cattle, compiled over several years. As

a result, large samples from a number of different breeds are available and analysis of this data will help in defining genetic relationships among all breeds. Additionally, blood typing reagents have been standardized internationally and this allows comparison of data from different countries.

In this chapter data collected by the Roslin Cattle Blood Typing Service on 19 British and 18 other European cattle breeds were used to assess breed relationships and test breed differentiation.

4.2 Materials and methods

4.2.1 Data

Animals included in the study were born between 1980-1995, and were located in the United Kingdom when blood typed. Pedigree information (parentage only) was available for about one third of samples, it was therefore not possible to select animals for the analysis that were known to be completely unrelated. Samples were generally submitted for blood typing because it is a requirement for breed society or herd book registration. For example, breed societies such as the Holstein-Friesian Society of Great Britain submit for parentage testing one in every thousand animals registered. It is likely, therefore, that the animals in the blood typing database represent a random sample of the pedigree populations in the United Kingdom. A total of 18,859 animals were included in the analysis, the number of animals sampled from each breed ranged from 46 White Park to 7,778 Holstein Friesians. Sample sizes for each individual breed are shown in table 4.1, and figure 4.1 shows the number of animals sampled from each breed compared with the UK breed census sizes given by the EAAP (1993) and MAFF (1996).

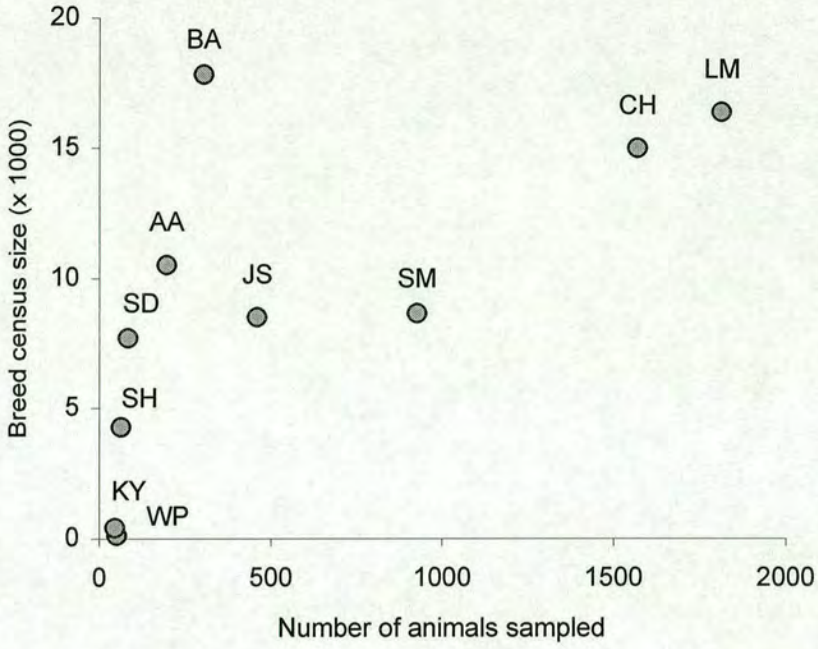

Figure 4.1 Relationship between the number of animals sampled from a breed for blood typing and the UK census size for that breed (as recorded by MAFF 1996).

In breeds which are routinely typed for pedigree verification (the breed society submits a random sample of animals presented for herd book registration each year, e.g. Limousin, Charolais, and Simmental Societies) the sample size is related to the census size. In other breeds the sample size may not be so well related. Figure 4.1 indicates that, for example, for the Blonde d'Aquitaine, the number of animals sampled is a smaller proportion of the total number of the breed in the United Kingdom, than has been sampled for other breeds.

4.2.2 Blood type markers

Seven red cell antigen systems A, B, C, F, L, S, Z and two serum proteins transferrin and albumin were the genetic markers used in this analysis. These nine markers are located on different chromosomes, and so provide a marker on about one in every three chromosomes. Details of the chromosomal location of the loci, and their individual antigenic factors (dominant loci) or alleles (co-dominant loci) are shown in chapter 2 (table 2.1). Each of the antigenic factors within the A, B, C, L, S and Z blood type

systems was considered to be a locus with two alleles, presence or absence of the factor, with absence of the factor being recessive (Neimann-Sorensen 1956). The F blood type system, albumin and transferrin are single co-dominant loci with either two alleles (F system, albumin) or four (transferrin).

4.2.3 Allele frequencies

Allele frequencies at the dominant loci (the antigenic factors within the A, B, C, L, S and Z systems) were estimated using an iterative allocation procedure (Ceppellini et al 1955; Weir 1996) which gives maximum likelihood estimates. Populations are assumed to be in Hardy-Weinberg equilibrium, and the observed frequency of each phenotype is divided into its constituent genotypes, according to the expected Hardy-Weinberg proportions. Gene frequency estimates are obtained by counting alleles, the phenotypes are then reallocated according to the new gene frequencies. The procedure is started with equal initial gene frequencies and repeated until the new gene frequency estimates converge. In the case of co-dominant loci (F system, transferrin and albumin) allele frequencies were estimated by direct gene counting.

4.2.4 Heterozygosities and number of alleles observed

Expected heterozygosities (under Hardy Weinberg equilibrium) were calculated for each of the populations as $1-\sum p^2$ averaged over all loci. The observed number of alleles in each breed were counted and expressed as a total number of alleles over all loci, since all loci (except transferrin) have only two alleles. Standard errors for the average heterozygosities and number of alleles were obtained by bootstrapping, with 200 bootstrap replicates, generated by resampling loci and individual phenotypes with replacement (see chapter 3 for details of the bootstrap procedure).

4.2.5 Effective population size

Under the infinite alleles model the expected heterozygosity at a locus is given by

$\overline{H}_e = \frac{4N_e v_0}{4N_e v_0 + 1}$ (Kimura 1983), where N_e is the effective population size and v_0 is the mutation rate per gene per generation for selectively neutral mutations. The average expected heterozygosity for each breed can therefore be used to obtain an estimate of the product $N_e v_0$. The average heterozygosity for each breed was compared with estimates of the effective population size obtained from census information (EAAP 1993; MAFF 1996), for breeds where census information was available. The effective population size (N_e) was estimated from the census numbers of males (N_m) and females (N_f) using $N_e = \frac{4N_m N_f}{N_m + N_f}$ (Falconer 1989).

4.2.6 Breed relationships and genetic distance

A principal component analysis (Genstat 1993) was carried out on the allele frequencies in order to summarise breed relationships. Genetic distances between breeds were calculated using the measure of Reynolds et al (1983):

$$D^2 = \frac{\sum_r \sum_i \left(p_{1ri} - p_{2ri}\right)^2}{2\sum_r \left(1 - \sum_i p_{1ri} p_{2ri}\right)}$$

where r = number of loci, i = number of alleles at the rth locus, p_{1ri} is the frequency of allele p_i in population 1 and p_{2ri} is the frequency of allele p_i in population 2.

This distance measure is based on Wright's F_{ST}, it assumes that populations have diverged due to drift alone, it also reflects the amount of gene flow between populations and is appropriate for the analysis of data sampled from a single species (Slatkin and Maddison 1990). Standard errors for the genetic distances were obtained by bootstrapping using 200 replicates, resampling both individual phenotypes and loci (for details of the bootstrap sampling procedure see chapter 3).

4.2.7 Graphical representation of genetic distance

A matrix of genetic distances can be difficult to interpret and summarise, particularly if there are a large number of populations involved. To represent the genetic distances graphically a multidimensional matrix must be reduced to two or three dimensions. Multidimensional scaling (Genstat 1993) is a method that produces a 'map' of the populations by reducing the multidimensional distance matrix to a set of co-ordinates representing the populations in just two or three dimensions. The co-ordinates selected to represent the populations are those that minimize the difference between the observed distances (the genetic distances) and the fitted distances (the distances between the co-ordinates). The goodness of fit of the distances between co-ordinates to the observed distances is measured by a stress statistic:

$$STRESS = \sqrt{\left\{ \frac{\sum \left(d_{ij} - \hat{d}_{ij} \right)^2}{\sum \hat{d}_{ij}^{\,2}} \right\}}$$

where $\hat{d}_{ij}$ = fitted distance and d_{ij} = observed distance. Small values of stress (between 0 and 0.1) are desirable (Manly 1986).

The Neighbour-joining method (Saitou and Nei 1987) was used to construct a dendrogram of breed relationships from the genetic distance matrix using the program PHYLIP (Felsenstein 1995). The robustness of the dendrogram was evaluated by bootstrapping (100 replicates), which was carried out by resampling loci, with linked loci sampled together.

4.2.8 Test of breed differentiation

A test of population differentiation can be carried out using a χ^2 test of allele frequencies (Workman and Niswander 1970). However, if sample sizes are small then the significance levels obtained from a χ^2 contingency test are not reliable. For

small sample sizes more reliable significance levels can be obtained using an exact permutation test (Hudson et al 1992). All data from the observed groups (populations or breeds) to be compared is pooled and then reassigned randomly and without replacement to the groups, keeping the number of observations per group the same as in the original data. The null hypothesis is that there is no differentiation among the populations. The number of possible permutations of the data matrix increases factorially with the number of breeds, alleles and total sample size, but the distribution can be approximated using Monte Carlo methods (Raymond and Rousset 1995).

Exact tests of population differentiation have been proposed by Hudson et al (1992), Roff and Bentzen (1989), and Raymond and Rousset (1995). In the first two papers an exact χ^2 was computed using permutation, in Raymond and Rousset (1995) the exact probability of observing the data under the null hypothesis was estimated using a Markov chain method. For both methods the test is carried out for each locus, and the observations which are randomized are the alleles observed in each population (genes within individuals).

In the analysis presented here the genetic distance was treated as the test statistic, since distance was the measure that was being used to assess the differences among breeds. Slatkin (1994) has previously suggested the use of F_{ST}, on which the genetic distance of Reynolds et al (1983) is based, as the test statistic when studying population differentiation. After each permutation of the data (genotypes were randomly assigned to breeds) the allele frequencies were re-estimated and the genetic distance among breeds recalculated. These new distances were then compared with the observed genetic distances, and the number of times the recalculated distance exceeded the observed distance counted to obtain p values (probability of obtaining the observed genetic distance by chance). Ten thousand permutations of the data were carried out.

4.3 Results

4.3.1 Heterozygosities, number of alleles and effective population sizes

Allele frequencies for the blood type loci (red cell antigen systems A, B, C, F, S, L, Z, and serum proteins transferrin and albumin) in all 37 breeds are shown in Appendix II. Average heterozygosities, observed number of alleles and their respective standard errors for all breeds are shown in table 4.1. The average heterozygosities ranged from 0.175 ($\pm$0.058) for the British White to 0.363 ($\pm$0.029) for the Romagnola. Observed number of alleles ranged between 53 for the Icelandic and 64 for the Charolais, Holstein-Friesian, Limousin and Simmental. Low heterozygosities and a small number of alleles were observed in the British White, Highland, Icelandic and White Park. This is consistent with the small population size of these breeds and suggests that the populations are closed (there is no migration of genes from other populations). Conversely, high heterozygosities and a larger number of alleles were observed in the Blonde d'Aquitaine, Romagnola, Limousin, Jersey and Guernsey. The French, Italian and Channel Island breeds were generally more heterozygous than other breeds.

A comparison of the average heterozygosities, with the effective population size estimated from census information (figure 4.2) indicates the Guernsey and Jersey are perhaps more heterozygous than would be expected from their census sizes. This may be because these breeds have been through a very recent population contraction, or there may be gene flow from other populations (e.g. imported animals or semen are being used in breeding programmes). In contrast, the Holstein-Friesian is less heterozygous than would be expected, which suggests that the number of animals used for breeding is much less than the census size.

Table 4.1 Number of animals sampled, average heterozygosity and number of alleles observed in each breed (standard errors in parentheses).

Country of origin	Breed	No. of animals sampled	Average heterozygosity	No. of alleles observed
Belgium	Belgian Blue	1242	0.26 (0.039)	64 (0.25)
Channel Islands	Guernsey	222	0.32 (0.031)	62 (0.34)
England	Jersey	460	0.30 (0.023)	63 (0.49)
	British White	71	0.18 (0.058)	59 (1.20)
	Gloucester	158	0.25 (0.036)	60 (0.67)
	Hereford	1034	0.24 (0.070)	61 (0.72)
	Poll Hereford	1079	0.23 (0.070)	62 (0.43)
	Lincoln Red	61	0.22 (0.040)	55 (0.84)
	Longhorn	66	0.21 (0.056)	61 (1.88)
	Shorthorn	64	0.27 (0.038)	59 (0.79)
	South Devon	84	0.23 (0.031)	61 (0.81)
	Sussex	64	0.22 (0.038)	61 (0.86)
	White Park	46	0.19 (0.032)	57 (1.02)
Scotland	Aberdeen Angus	197	0.24 (0.043)	62 (0.69)
	Ayrshire	154	0.23 (0.052)	60 (0.74)
	Belted Galloway	73	0.21 (0.041)	59 (0.88)
	Galloway	64	0.22 (0.040)	62 (0.89)
	Highland	57	0.19 (0.039)	57 (0.71)
	Murray Grey	61	0.22 (0.039)	58 (0.51)
Wales	Welsh Black	65	0.26 (0.047)	61 (1.15)
France	Blonde d'Aquitaine	304	0.34 (0.039)	63 (0.62)
	Charolais	1569	0.29 (0.049)	64 (0.00)
	Limousin	1813	0.32 (0.024)	64 (0.00)
	Maine-Anjou	86	0.27 (0.037)	63 (0.97)
	Salers	85	0.28 (0.045)	62 (0.37)
Germany	Gelbvieh	99	0.25 (0.058)	62 (1.09)
Iceland	Icelandic	82	0.19 (0.033)	53 (0.33)
Ireland	Dexter	247	0.28 (0.033)	63 (0.55)
	Kerry	52	0.25 (0.035)	57 (0.45)
Italy	Chianina	66	0.28 (0.059)	59 (0.76)
	Marchigiana	77	0.29 (0.038)	62 (0.82)
	Piemontese	69	0.29 (0.035)	62 (0.33)
	Romagnola	78	0.36 (0.029)	64 (0.26)
Netherlands	Holstein Friesian	7778	0.24 (0.036)	64 (0.00)
	Meuse Rhine Yssel	99	0.23 (0.049)	58 (0.67)
Switzerland	Brown Swiss	68	0.29 (0.027)	62 (0.78)
	Simmental	928	0.24 (0.045)	64 (0.00)

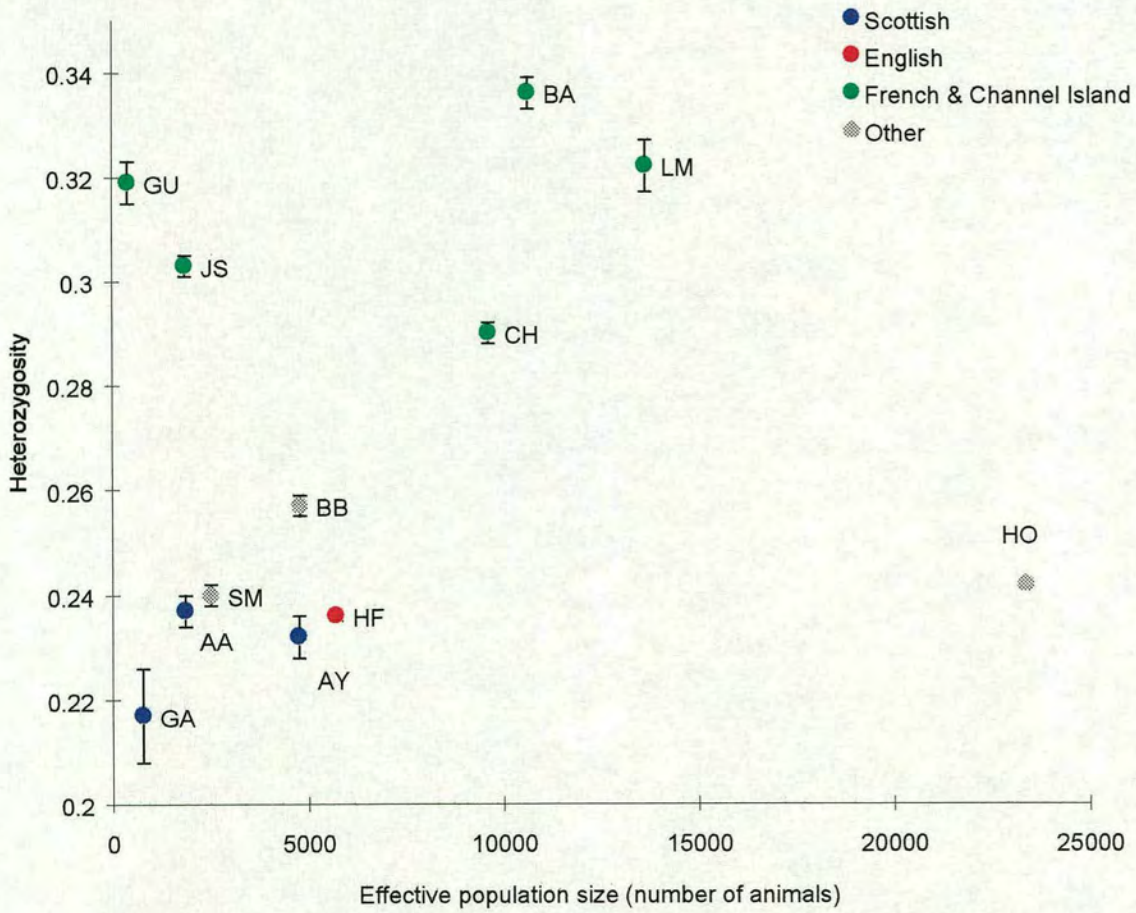

Figure 4.2 Relationship between average heterozygosity and effective population size estimated from census size

4.3.2 Test of breed differentiation

The permutation test showed all breeds to be significantly different from one another ($p < 0.0001$), even breeds that might be expected to be closely related (e.g. Dexter and Kerry, or Galloway and Belted Galloway). Buys and Chiperzak (1992) also found that the Dexter and Kerry differed significantly in blood type gene frequencies, and concluded they should be considered as separate breeds. The permutation test results suggest that there has been very limited gene flow among modern day pedigree populations which has resulted in significant differentiation among breeds.

4.3.3 Relationships and genetic distances among breeds

Principal component scores for the first three components from the analysis of allele frequencies are plotted in figure 4.3. The third component or dimension was represented by the diameter of the points, with points that are distant in the third dimension having a smaller diameter than those that are closer to the viewer. The analysis indicated a grouping of French, Italian and Channel Island breeds (Blonde d'Aquitaine, Charolais, Limousin, Salers, Chianina, Marchigiana, Piemontese, Romagnola, Guernsey, Jersey) together with the Simmental and Gelbvieh. The other breeds formed a larger grouping, the Scottish breeds (Galloway, Highland, Murray Grey, Belted Galloway, Aberdeen Angus) forming a sub-group with the British White and White Park within this larger group. Only 48% of the variance was accounted for by the first three dimensions of the principal component analysis. Nine components were required to account for 80% of the variance.

Bootstrap means and standard errors of the genetic distances among breeds are shown in table 4.2. Mean distances ranged from 0.011 (Hereford-Poll Hereford) and 0.017 (Ayrshire–Meuse Rhine Yssel) to 0.309 (Icelandic-Jersey) and 0.292 (Icelandic-Longhorn; White Park-Chianina). Standard errors ranged from 0.005

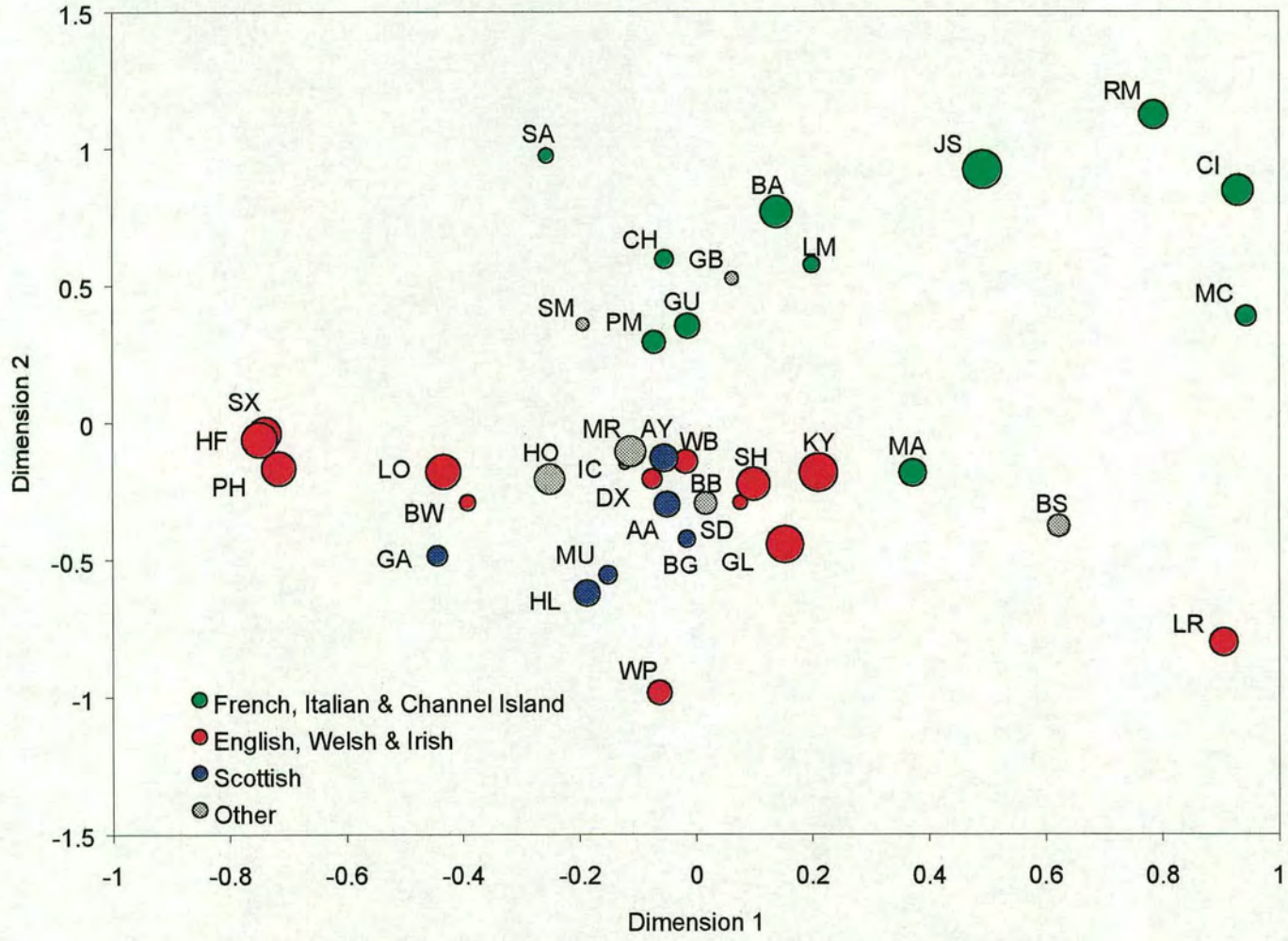

Figure 4.3 Principal component analysis of allele frequencies on 37 European cattle breeds. Three components are plotted, with the third represented by the diameter of the point, so points that are distant from the viewer are smaller than those that are closer.

Table 4.2 Genetic distances (below diagonal) and standard errors (above diagonal). Standard errors were estimated using bootstrapping by resampling loci and individual genotypes, with linked loci sampled together.

	AA	AY	BB	BG	BA	BW	BS	CH	CI	DX	HO	GA	GB
AA		0.020	0.016	0.015	0.022	0.030	0.043	0.022	0.044	0.016	0.026	0.023	0.041
AY	0.045		0.012	0.017	0.016	0.024	0.053	0.027	0.040	0.021	0.019	0.024	0.039
BB	0.042	0.031		0.016	0.033	0.024	0.033	0.030	0.039	0.016	0.013	0.015	0.042
BG	0.033	0.054	0.035		0.045	0.026	0.063	0.043	0.045	0.021	0.034	0.021	0.062
BA	0.101	0.085	0.107	0.137		0.029	0.033	0.013	0.036	0.030	0.030	0.039	0.027
BW	0.079	0.068	0.083	0.067	0.148		0.086	0.024	0.047	0.023	0.039	0.036	0.057
BS	0.074	0.110	0.081	0.102	0.129	0.160		0.039	0.045	0.029	0.041	0.047	0.042
CH	0.104	0.088	0.106	0.125	0.049	0.112	0.142		0.047	0.031	0.025	0.043	0.029
CI	0.176	0.149	0.176	0.204	0.093	0.241	0.188	0.126		0.034	0.040	0.038	0.065
DX	0.043	0.046	0.042	0.052	0.102	0.083	0.079	0.089	0.168		0.015	0.020	0.042
HO	0.046	0.044	0.033	0.056	0.105	0.094	0.110	0.119	0.206	0.055		0.017	0.075
GA	0.054	0.067	0.044	0.048	0.156	0.096	0.133	0.148	0.254	0.049	0.040		0.070
GB	0.110	0.107	0.126	0.138	0.083	0.149	0.147	0.089	0.120	0.109	0.142	0.175	
GL	0.142	0.133	0.139	0.149	0.183	0.176	0.193	0.208	0.248	0.141	0.140	0.153	0.233
GU	0.072	0.066	0.067	0.091	0.069	0.116	0.125	0.080	0.124	0.061	0.076	0.089	0.092
HF	0.127	0.105	0.130	0.151	0.119	0.135	0.189	0.138	0.232	0.109	0.129	0.128	0.168
PH	0.104	0.085	0.113	0.127	0.122	0.115	0.182	0.132	0.236	0.090	0.111	0.106	0.173
HL	0.074	0.080	0.069	0.052	0.168	0.120	0.138	0.159	0.252	0.084	0.097	0.090	0.184
IC	0.155	0.169	0.153	0.153	0.217	0.186	0.213	0.184	0.230	0.122	0.192	0.182	0.125
JS	0.198	0.164	0.173	0.230	0.086	0.252	0.194	0.132	0.146	0.180	0.188	0.239	0.162
KY	0.087	0.079	0.108	0.144	0.101	0.156	0.136	0.133	0.187	0.107	0.098	0.152	0.175
LM	0.098	0.079	0.080	0.110	0.041	0.123	0.112	0.049	0.130	0.088	0.099	0.127	0.104
LO	0.148	0.177	0.190	0.210	0.168	0.191	0.177	0.158	0.289	0.164	0.180	0.204	0.214
LR	0.126	0.137	0.114	0.126	0.199	0.185	0.108	0.200	0.227	0.138	0.170	0.188	0.222
MA	0.059	0.040	0.054	0.065	0.095	0.099	0.091	0.092	0.152	0.069	0.060	0.103	0.140
MC	0.152	0.143	0.166	0.170	0.112	0.201	0.161	0.113	0.064	0.136	0.189	0.230	0.147
MR	0.042	0.017	0.031	0.045	0.085	0.080	0.122	0.096	0.155	0.051	0.036	0.061	0.114
MU	0.048	0.074	0.065	0.038	0.152	0.087	0.110	0.139	0.239	0.055	0.063	0.061	0.158
PM	0.044	0.050	0.056	0.069	0.051	0.093	0.099	0.066	0.116	0.040	0.050	0.078	0.071
RM	0.170	0.166	0.169	0.205	0.074	0.243	0.163	0.129	0.088	0.167	0.180	0.230	0.116
SA	0.140	0.134	0.146	0.177	0.053	0.178	0.193	0.047	0.148	0.136	0.140	0.177	0.108
SH	0.055	0.053	0.060	0.069	0.091	0.093	0.086	0.092	0.168	0.045	0.078	0.096	0.128
SM	0.069	0.064	0.076	0.082	0.075	0.098	0.124	0.057	0.150	0.064	0.079	0.099	0.061
SD	0.084	0.092	0.079	0.080	0.140	0.090	0.111	0.138	0.213	0.107	0.133	0.132	0.139
SX	0.140	0.122	0.133	0.168	0.166	0.159	0.225	0.162	0.259	0.109	0.100	0.106	0.218
WB	0.057	0.037	0.061	0.069	0.100	0.071	0.114	0.090	0.163	0.049	0.067	0.087	0.091
WP	0.079	0.110	0.074	0.085	0.213	0.162	0.107	0.215	0.292	0.097	0.111	0.076	0.235

Table 4.2 Genetic distances (below diagonal) and standard errors (above diagonal). Standard errors were estimated using bootstrapping by resampling loci and individual genotypes, with linked loci sampled together (continued).

	GL	GU	HF	PH	HL	IC	JS	KY	LM	LO	LR	MA
AA	0.045	0.021	0.036	0.037	0.029	0.056	0.045	0.043	0.017	0.065	0.050	0.022
AY	0.028	0.024	0.028	0.023	0.027	0.045	0.033	0.034	0.018	0.046	0.056	0.016
BB	0.033	0.019	0.025	0.028	0.031	0.048	0.044	0.054	0.018	0.071	0.045	0.024
BG	0.047	0.028	0.037	0.038	0.036	0.070	0.061	0.066	0.035	0.085	0.062	0.024
BA	0.061	0.014	0.026	0.020	0.058	0.062	0.021	0.031	0.012	0.056	0.062	0.035
BW	0.052	0.030	0.043	0.049	0.038	0.077	0.044	0.051	0.033	0.064	0.091	0.036
BS	0.057	0.042	0.063	0.068	0.066	0.047	0.060	0.058	0.032	0.076	0.027	0.037
CH	0.048	0.020	0.023	0.025	0.048	0.057	0.029	0.035	0.012	0.042	0.041	0.030
CI	0.076	0.043	0.066	0.070	0.058	0.048	0.046	0.043	0.038	0.060	0.060	0.050
DX	0.028	0.016	0.018	0.022	0.027	0.036	0.049	0.042	0.024	0.068	0.044	0.022
HO	0.034	0.011	0.031	0.028	0.050	0.073	0.043	0.044	0.023	0.072	0.076	0.020
GA	0.031	0.026	0.024	0.028	0.034	0.079	0.056	0.065	0.036	0.090	0.069	0.032
GB	0.082	0.027	0.050	0.052	0.063	0.060	0.040	0.051	0.029	0.051	0.061	0.054
GL		0.035	0.049	0.064	0.061	0.087	0.092	0.041	0.048	0.058	0.052	0.050
GU	0.113		0.044	0.046	0.032	0.044	0.028	0.028	0.020	0.054	0.038	0.025
HF	0.169	0.113		0.005	0.054	0.052	0.040	0.032	0.028	0.047	0.061	0.035
PH	0.141	0.105	0.011		0.038	0.053	0.046	0.045	0.025	0.049	0.071	0.040
HL	0.169	0.142	0.144	0.113		0.065	0.072	0.061	0.045	0.083	0.096	0.035
IC	0.261	0.130	0.228	0.222	0.197		0.071	0.084	0.066	0.089	0.064	0.060
JS	0.220	0.107	0.224	0.227	0.259	0.309		0.053	0.026	0.075	0.073	0.044
KY	0.136	0.102	0.149	0.128	0.162	0.253	0.171		0.027	0.042	0.057	0.031
LM	0.195	0.061	0.158	0.153	0.166	0.205	0.105	0.123		0.043	0.053	0.028
LO	0.225	0.158	0.155	0.148	0.223	0.292	0.252	0.139	0.195		0.057	0.045
LR	0.199	0.173	0.241	0.223	0.145	0.242	0.256	0.144	0.172	0.256		0.040
MA	0.125	0.088	0.151	0.130	0.099	0.198	0.178	0.061	0.078	0.174	0.106	
MC	0.231	0.130	0.228	0.220	0.204	0.184	0.199	0.166	0.140	0.253	0.174	0.121
MR	0.135	0.060	0.100	0.080	0.065	0.151	0.173	0.082	0.088	0.196	0.144	0.050
MU	0.139	0.102	0.147	0.117	0.054	0.165	0.258	0.120	0.130	0.182	0.137	0.055
PM	0.149	0.045	0.116	0.105	0.113	0.136	0.129	0.114	0.064	0.171	0.179	0.083
RM	0.262	0.115	0.222	0.231	0.262	0.261	0.118	0.181	0.093	0.250	0.247	0.163
SA	0.261	0.084	0.178	0.180	0.225	0.215	0.135	0.168	0.055	0.209	0.274	0.140
SH	0.101	0.068	0.104	0.085	0.092	0.171	0.170	0.063	0.090	0.137	0.114	0.055
SM	0.221	0.069	0.140	0.138	0.140	0.142	0.159	0.161	0.069	0.209	0.211	0.099
SD	0.197	0.116	0.168	0.158	0.124	0.192	0.226	0.168	0.106	0.184	0.136	0.110
SX	0.176	0.100	0.117	0.106	0.205	0.237	0.235	0.136	0.156	0.195	0.266	0.156
WB	0.098	0.059	0.102	0.086	0.103	0.131	0.175	0.091	0.105	0.138	0.145	0.069
WP	0.209	0.177	0.188	0.165	0.081	0.245	0.300	0.181	0.195	0.229	0.141	0.128

Table 4.2 Genetic distances (below diagonal) and standard errors (above diagonal). Standard errors were estimated using bootstrapping by resampling loci and individual genotypes, with linked loci sampled together (continued).

	MC	MR	MU	PM	RM	SA	SH	SM	SD	SX	WB	WP
AA	0.060	0.021	0.026	0.016	0.065	0.040	0.029	0.018	0.042	0.058	0.033	0.015
AY	0.059	0.013	0.030	0.024	0.041	0.035	0.025	0.018	0.047	0.035	0.028	0.035
BB	0.053	0.020	0.025	0.026	0.056	0.046	0.021	0.017	0.052	0.042	0.038	0.027
BG	0.066	0.021	0.023	0.038	0.078	0.062	0.033	0.040	0.048	0.063	0.046	0.051
BA	0.036	0.026	0.059	0.019	0.025	0.018	0.031	0.018	0.058	0.043	0.040	0.067
BW	0.065	0.027	0.023	0.027	0.063	0.045	0.054	0.036	0.033	0.053	0.022	0.069
BS	0.049	0.057	0.054	0.038	0.061	0.042	0.034	0.060	0.026	0.061	0.062	0.040
CH	0.037	0.033	0.049	0.021	0.052	0.020	0.037	0.021	0.048	0.036	0.036	0.064
CI	0.038	0.042	0.053	0.043	0.032	0.064	0.048	0.047	0.050	0.060	0.037	0.049
DX	0.048	0.028	0.029	0.014	0.067	0.048	0.016	0.018	0.050	0.042	0.021	0.041
HO	0.053	0.022	0.040	0.027	0.059	0.043	0.032	0.028	0.105	0.041	0.041	0.050
GA	0.055	0.029	0.031	0.036	0.066	0.064	0.031	0.040	0.089	0.049	0.031	0.036
GB	0.057	0.038	0.084	0.034	0.035	0.044	0.051	0.037	0.068	0.075	0.036	0.058
GL	0.086	0.037	0.050	0.047	0.089	0.074	0.032	0.056	0.060	0.044	0.026	0.066
GU	0.043	0.027	0.034	0.015	0.046	0.037	0.028	0.021	0.043	0.026	0.022	0.051
HF	0.063	0.033	0.046	0.033	0.055	0.039	0.036	0.027	0.053	0.044	0.034	0.051
PH	0.074	0.033	0.035	0.034	0.059	0.042	0.045	0.024	0.044	0.033	0.035	0.051
HL	0.071	0.032	0.033	0.048	0.077	0.077	0.030	0.060	0.046	0.075	0.048	0.048
IC	0.050	0.048	0.085	0.051	0.084	0.099	0.066	0.065	0.093	0.094	0.050	0.081
JS	0.044	0.046	0.072	0.037	0.032	0.048	0.042	0.052	0.059	0.043	0.040	0.079
KY	0.059	0.042	0.052	0.037	0.064	0.049	0.025	0.041	0.068	0.040	0.042	0.093
LM	0.029	0.021	0.053	0.020	0.038	0.018	0.025	0.014	0.046	0.045	0.027	0.049
LO	0.059	0.053	0.072	0.051	0.082	0.062	0.045	0.064	0.058	0.073	0.050	0.128
LR	0.067	0.049	0.063	0.061	0.106	0.067	0.030	0.060	0.046	0.076	0.049	0.064
MA	0.056	0.026	0.031	0.038	0.070	0.048	0.020	0.032	0.061	0.053	0.042	0.048
MC		0.050	0.066	0.046	0.065	0.064	0.061	0.044	0.068	0.056	0.058	0.075
MR	0.140		0.020	0.023	0.044	0.042	0.020	0.036	0.067	0.043	0.028	0.042
MU	0.188	0.064		0.047	0.088	0.082	0.023	0.067	0.070	0.074	0.041	0.041
PM	0.115	0.048	0.097		0.047	0.031	0.024	0.017	0.063	0.046	0.028	0.062
RM	0.134	0.162	0.229	0.114		0.053	0.069	0.049	0.096	0.065	0.050	0.070
SA	0.163	0.134	0.183	0.089	0.127		0.051	0.031	0.083	0.051	0.051	0.080
SH	0.145	0.051	0.066	0.071	0.166	0.161		0.032	0.033	0.043	0.020	0.051
SM	0.149	0.069	0.115	0.040	0.143	0.067	0.109		0.062	0.054	0.043	0.047
SD	0.205	0.111	0.121	0.104	0.207	0.191	0.089	0.130		0.111	0.052	0.062
SX	0.253	0.116	0.148	0.129	0.238	0.184	0.122	0.176	0.217		0.043	0.089
WB	0.140	0.051	0.080	0.058	0.168	0.149	0.051	0.080	0.104	0.122		0.066
WP	0.248	0.121	0.094	0.152	0.286	0.268	0.137	0.182	0.137	0.227	0.145	

(Hereford-Poll Hereford) to 0.105 (Holstein-Friesian-South Devon) and 0.128 (Longhorn-White Park). Representation of the genetic distance matrix using multidimensional scaling (figure 4.4) gave a similar pattern to that obtained from the principal component analysis. The French, Italian and Channel Island breeds were clustered with the Gelbvieh and Simmental. The other breeds formed a larger cluster with two outlying branches, one consisting of the Sussex, Hereford and Longhorn, the other of the Maine-Anjou, Gloucester, Kerry, Brown Swiss and Lincoln Red. The Icelandic, British White and White Park were distant from other breeds in the third dimension. The value of the stress statistic was 0.12 indicating that there was some discrepancy between the fitted and observed distances. Figure 4.5 shows the decline in the stress statistic as the number of dimensions permitted for the multidimensional scaling solution is increased. A decrease in the stress statistic could be obtained by increasing the number of dimensions of the solution.

The neighbour joining tree (figure 4.6) indicated two major breed groups, which could be divided into four smaller sub-groups. The first sub-group consisted of the French (Blonde d'Aquitaine, Charolais, Limousin, Salers) and Italian (Chianina, Marchigiana, Romagnola) breeds with the Jersey. The second sub-group consisted of the Kerry, Gloucester, Shorthorn, Maine-Anjou, Ayrshire, Meuse Rhine Yssel, Brown Swiss, Lincoln Red and Welsh Black. The third was made up of the Longhorn, Sussex and Hereford. Finally, the fourth sub-group consisted of the Aberdeen Angus, White Park, British White, Belted Galloway, Murray Grey, Highland, Belgian Blue, Galloway, Holstein Friesian and South Devon. The Piemontese, Guernsey and Dexter were intermediate between the two major groups, and the Icelandic, Gelbvieh and Simmental were outlying to all other breeds. Bootstrap values were generally low, however, indicating that the relationships inferred from this small portion of the genome (31 loci on 9 chromosomes) are not robust.

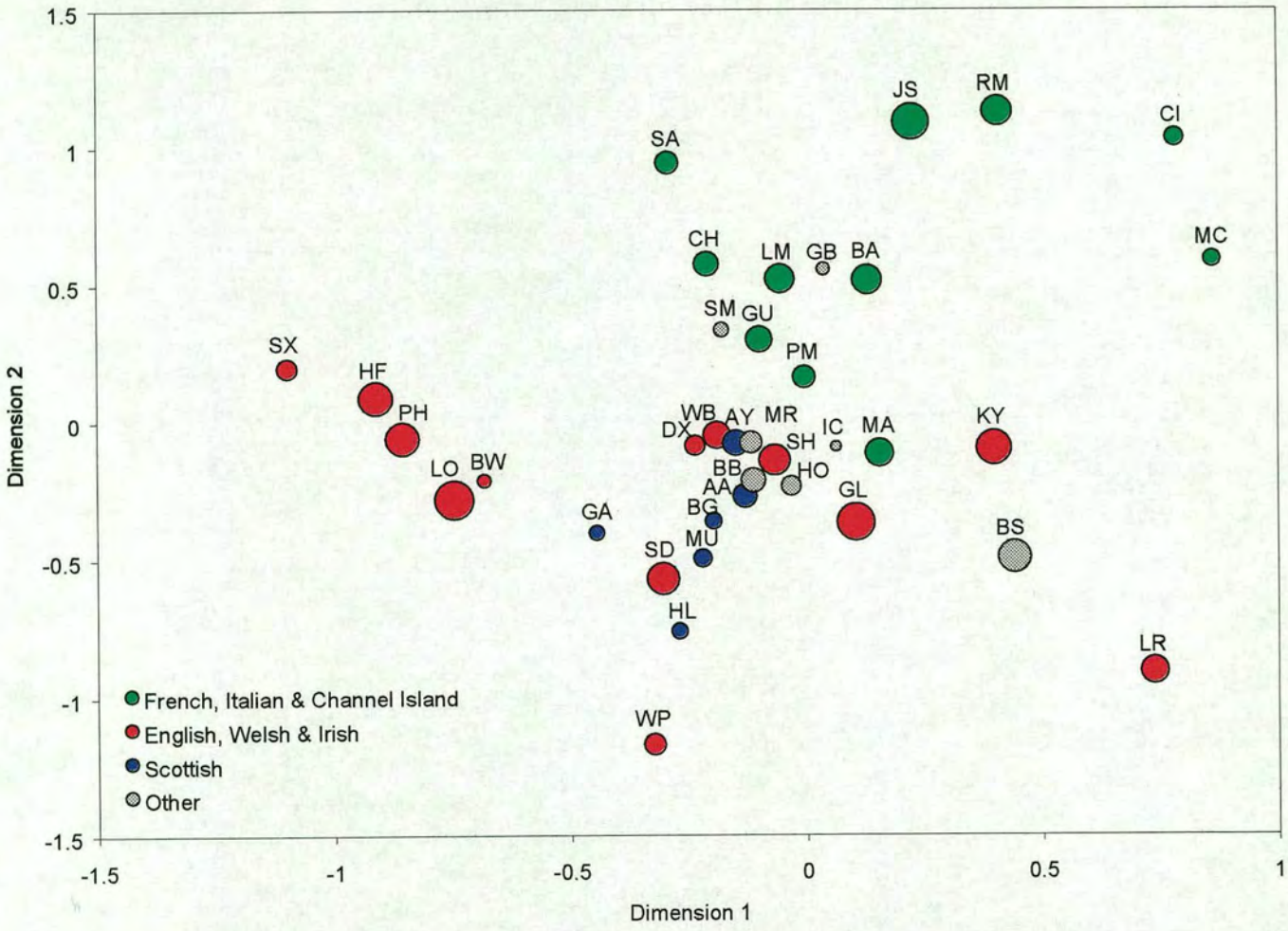

Figure 4.4a. Three-dimensional representation of the genetic distances among 37 European cattle breeds obtained using multidimensional scaling. Distance in the third dimension is represented by the diameter of the point, so points that are distant from the viewer are smaller than those that are closer. Breeds are colour coded according to their geographic region of origin.

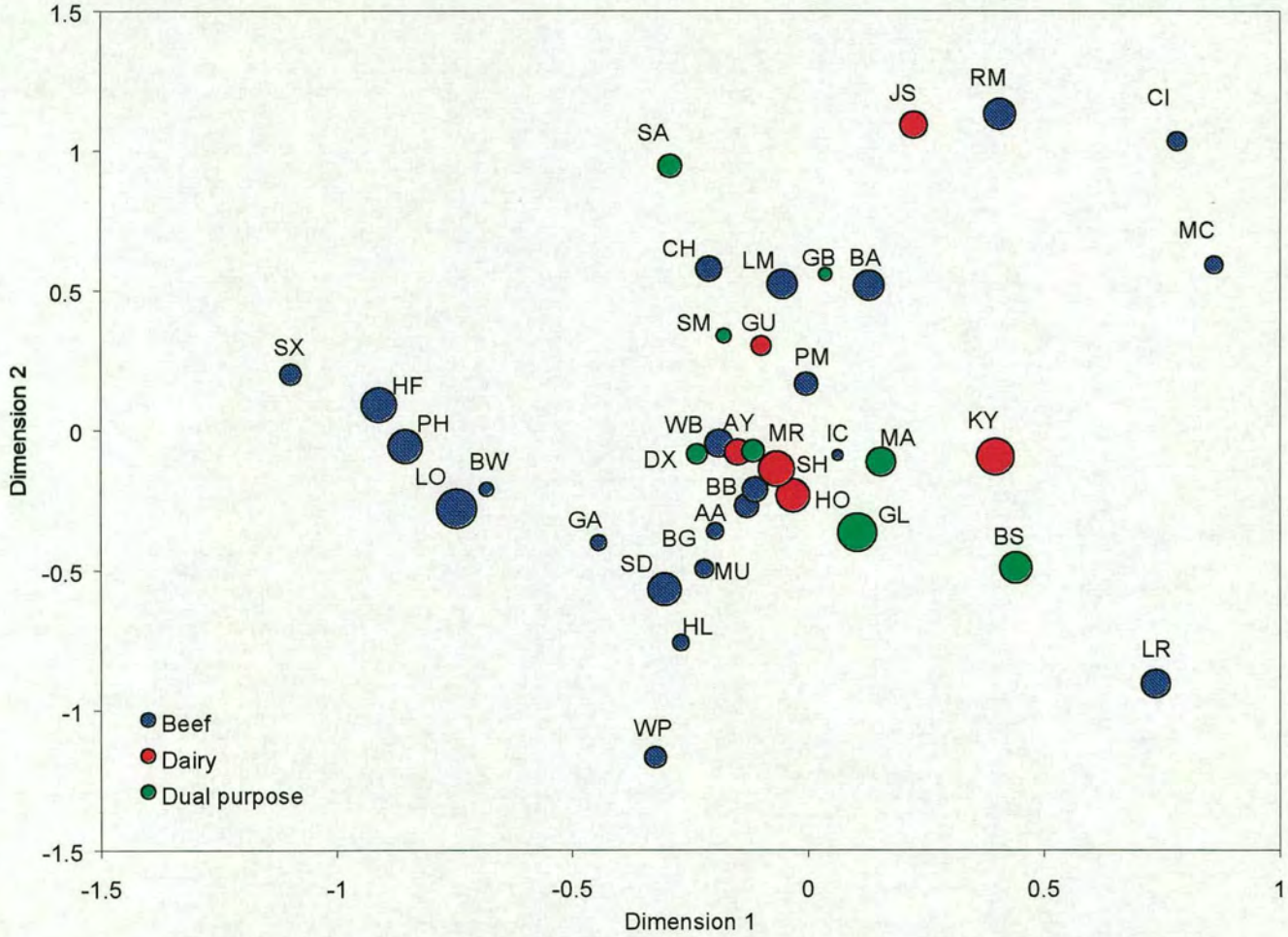

Figure 4.4b Three dimensional representation of the genetic distances among 37 European cattle breeds obtained using multidimensional scaling. Distance in the third dimension is represented by the diameter of the point, so points that are distant from the viewer are smaller than those that are closer. Breeds are colour coded according to the purpose for which they have been selected.

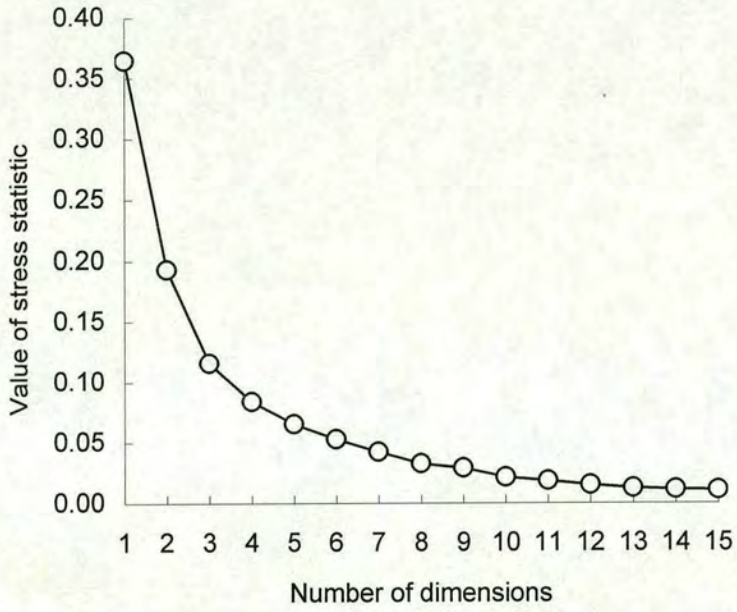

Figure 4.5 Decline in the stress statistic as the number of dimensions for the multidimensional scaling solution is increased.

Comparison of the multidimensional scaling and principal component plots with the neighbour joining tree highlights some discrepancies between the different representations of the distance matrix. The neighbour joining tree shows the breeds falling into four sub-groups. Both the principal component analysis and the multidimensional scaling confirmed the first group (French, Italian and Channel Island breeds), but the other groups were not so clear. The tree indicates that the Piemontese, Guernsey and Dexter are intermediate between two major breed groups. However, the multidimensional scaling and principal component analysis suggest that the Piemontese and Guernsey are part of a group of French, Italian and Channel Island breeds, and the Dexter part of a mainland British group. The tree also indicates that the Icelandic, Gelbvieh and Simmental are outlying breeds, whereas the multidimensional scaling and principal component analysis suggest that the Gelbvieh and Simmental group with the French, Italian and Channel Island breeds.

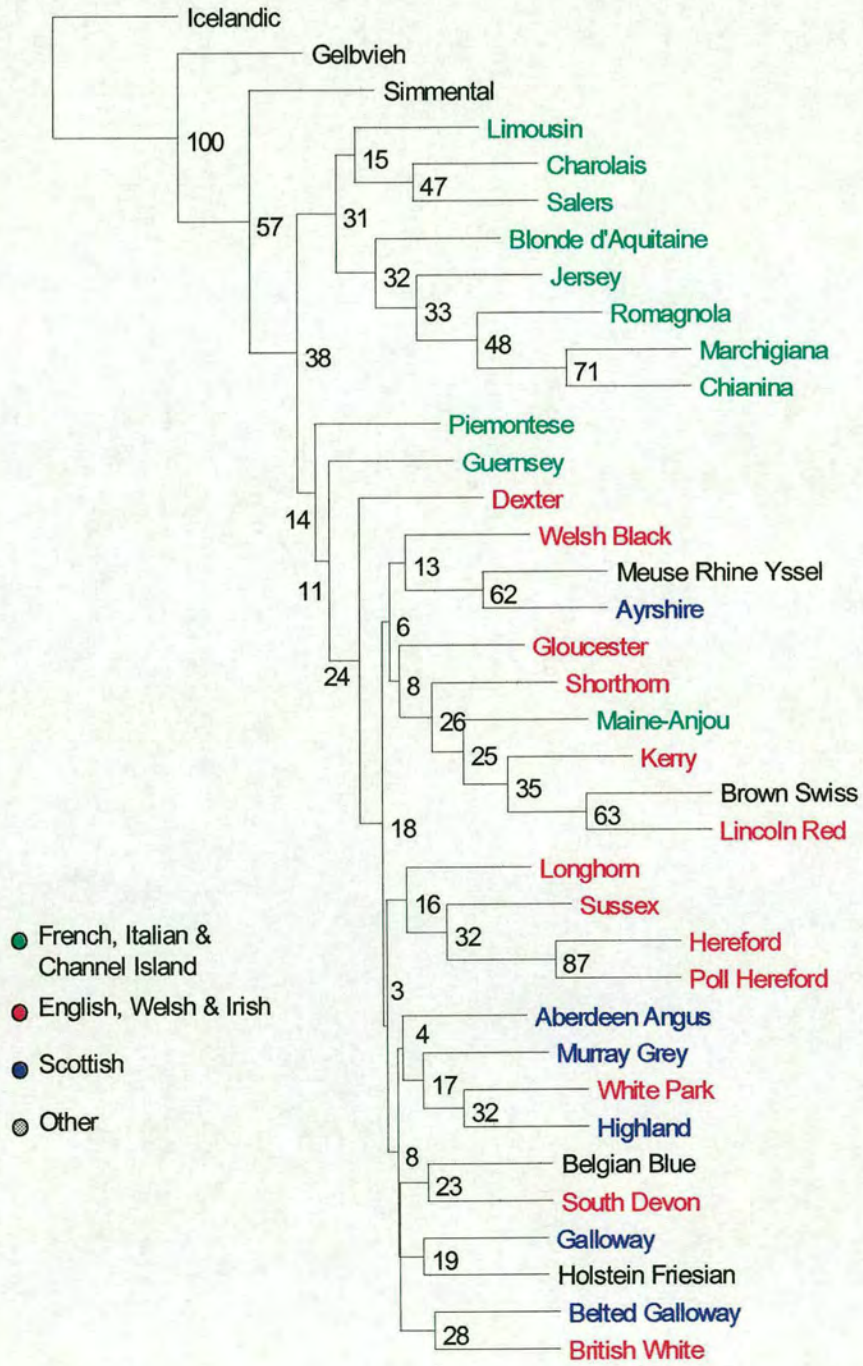

Figure 4.6 Unrooted neighbour joining tree showing the relationships among 37 European cattle breeds. Bootstrap proportions (number of times node was observed in 100 replicates) are shown at each node. Bootstrapping was carried out by resampling loci, with linked loci sampled together.

4.3.4 Relationship between genetic distances among breeds and the geographic distances among places of breed origin

The relationship between genetic distance and geographic distance was further investigated by comparing the estimated genetic distances (Reynolds et al 1983) with geographic distances between breed origins estimated from the scale map in Porter (1991). Geographic distances were measured on the map "as the crow flies" and converted from millimetres to kilometres. Figure 4.7 shows the geographic distances between breed origins plotted against the genetic distances between breeds.

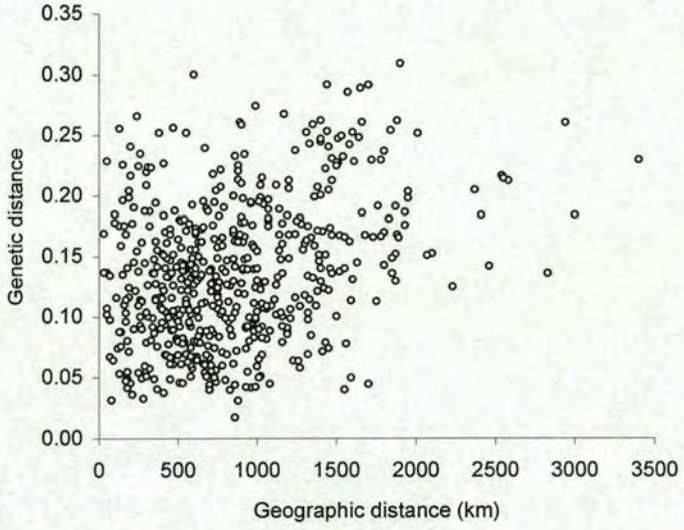

Figure 4.7 Relationship between geographic distance between breed origins and genetic distance between breeds.

The correlation between the two distance matrices (geographic and genetic) was 0.18. This correlation was found to be not significantly different from zero at the 5% level of significance ($p = 0.052$) using a Mantel test (Mantel 1967; Manly 1986). In section 4.3.3 it was suggested that breeds could be placed into groups based on their geographic origin. Four groups were defined; the first consisting of French, Italian and Channel Island breeds, the second English, Welsh and Irish breeds, the third Scottish breeds and the fourth "other" breeds (including the Holstein-Friesian, Belgian Blue, Gelbvieh and Simmental). These categories were defined by the pattern of breed groupings obtained using principal component analysis and

multidimensional scaling. As a further test of the relationship between genetic and geographic grouping the average genetic distances between breeds within each of these groups and between breeds in different groups were computed, and compared to the geographic distances within and between groups. Table 4.3 shows the average genetic and geographic distances within and between the different breed groups.

Table 4.3 Genetic and geographic distances between breed groups. Wtthin group distances on the diagonal, between group distances off the diagonal. In each cell the top figure is the genetic distance and the bottom figure the geographic distance (km). Standard errors are shown in parentheses.

Breed group	French, Italian & Channel Island	English, Welsh & Irish	Scottish	Other
French, Italian & Channel Island	0.104 (0.04) 580 (460)			
English, Welsh & Irish	0.167 (0.06) 1000 (456)	0.141 (0.05) 260 (186)		
Scottish	0.147 (0.06) 1360 (408)	0.113 (0.04) 470 (170)	0.054 (0.02) 100 (102)	
Other	0.131 (0.06) 1000 (678)	0.142 (0.06) 930 (413)	0.097 (0.05) 1080 (233)	0.111 (0.05) 680 (850)

Figure 4.8 illustrates the relationship between the average genetic and geographic distances both between and within groups. In general, the within group distances are smaller than those between groups and there is a positive relationship between the genetic and geographic distances. The fact that the within group distances are smaller than those between groups supports the view that the breeds can be clustered into these geographic groupings. The non-significant correlation between individual genetic distances between breeds and geographic distances between their places of origin is perhaps not surprising. A number of different factors will have affected the genetic history of breeds. Geographic isolation is probably not just due to distance between populations but also due to barriers such as the presence of mountainous regions or the sea having caused breeds to be isolated. The migration patterns of human populations may also have been a factor in determining the type of cattle that

were introduced into different regions. Breeding practices will also have influenced the genetic make-up of breeds and the effects of admixture and selection must be taken into account.

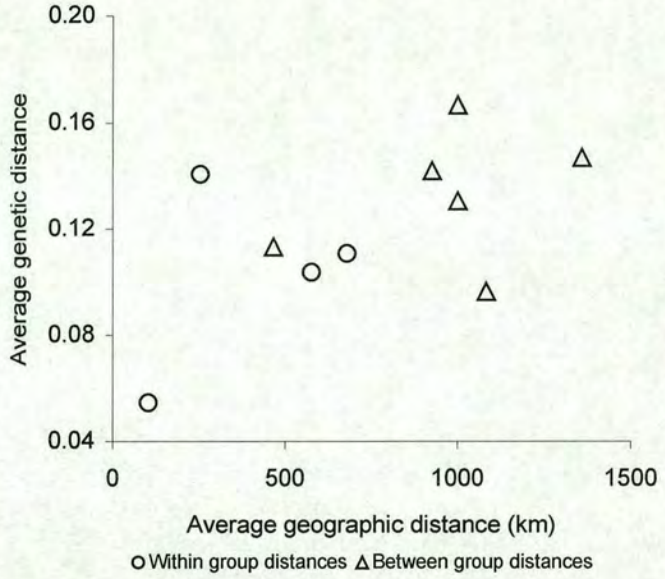

Figure 4.8 Relationship between the average genetic distance and average geographic distance both within and between breed groups.

4.4 Discussion

Three different methods of graphically representing breed relationships were presented; principal component analysis, multidimensional scaling and a neighbour joining tree. Although the principal component analysis was carried out on gene frequencies and the multidimensional scaling on the genetic distances the resulting plots of breed relationships were in good agreement. The neighbour joining tree, however, produced slightly different results. The overall pattern of breed relationships in the tree was similar to that observed with the other two methods, but the placement of some breeds was anomalous (e.g. the Holstein-Friesian was grouped with the Galloway). All three methods indicated two major breed groups; the French, Italian, and Channel Island breeds together with the Simmental and

Gelbvieh, (a 'continental' group of 12 breeds) and a larger group consisting mainly of the British and North European breeds (25 breeds).

Grosclaude et al (1990) found that breeds from central and south-west France, including the Charolais, Blonde d'Aquitaine, Salers and Limousin, formed a group that was separated from breeds from the north, including the Maine-Anjou and Shorthorn. The results presented in this thesis confirm the separation of the Maine-Anjou from these other French breeds, placing it in the group of British and North European breeds. This is in concordance with the breed's ancestry, as it originated from the crossing of Shorthorns with a local French breed the Mancelle (Porter 1991).

The Italian breeds (Piemontese, Chianina, Marchigiana and Romagnola) were also found to group with the breeds from central and south-west France and the Channel Islands, although the Chianina, Marchigiana and Romagnola were more distant than the Piemontese. This separation of the Piemontese from the other three Italian breeds has been previously noted by Ciampolini et al (1995) and Astolfi et al (1983). The Jersey, Salers, Romagnola Chianina and Marchigiana were the most outlying breeds in the group. Medjugorac et al (1994) found the Jersey to be distant from Balkan and other central European breeds. Historically, the Channel Island breeds are known to have a common ancestry with breeds from France (Porter 1991) and belong to the group of breeds which originated in the Mediterranean and south-west Europe (Astolfi et al 1983). The Simmental and Gelbvieh were also found on the periphery of this group of French and Italian breeds. Both the Simmental and Gelbvieh are believed to have a common ancestor, the Bernese, and Charolais was used to improve local breeds during the development of the Gelbvieh (Porter 1991).

The Belgian Blue, Holstein Friesian and Meuse Rhine Yssel (from Belgium and the Netherlands respectively) were shown to be genetically close to the Shorthorn and Ayrshire. In the case of the Belgian Blue this is probably due to shared ancestry, the breed originated from crosses between Dutch Black Pied (Friesian) and Shorthorn.

The proximity of the Shorthorn and Ayrshire to the Holstein-Friesian and Meuse Rhine Yssel may be a result of more recent crossing. In the United Kingdom the Dairy Shorthorn has been crossed with the Red Holstein and Meuse Rhine Yssel, and in the United States with Red Holstein, in a programme of genetic upgrading (Porter 1991). A similar situation now exists with the Ayrshire breed, where Red Holstein sires are being used in Ayrshire herds. In the past the Ayrshire has also been influenced by Shorthorn blood.

The outlying or most distant breeds in the British and North European group were the Sussex, Hereford, White Park, Lincoln Red and Icelandic. Royle (1983) previously found the White Park to be distant from eight other British breeds based on an analysis of blood groups. The results presented here show the nearest related breeds to the White Park are the Scottish breeds.

The Brown Swiss was also, perhaps unexpectedly, within the group of British and North European breeds, although at some distance from the other breeds within the group. Brown Swiss animals included in this study, although located in the United Kingdom, were of American breeding. American Brown Swiss are descended from 135 Swiss Brown (Braunvieh) cattle imported to the USA in 1869 (Porter 1991). Medjugorac et al (1994) also found that the American Brown Swiss was distinct from other breeds, including populations of the German Braunvieh and Slovenian Braunvieh.

The apparent taxonomic distinctiveness of a breed may not necessarily mean that it carries genes that are adaptively unique, as forces other than selection may have been operating. For example, random drift can affect the genetic distances among populations. Takezaki and Nei (1996) have shown that when a population goes through a bottleneck there are two effects: i) genetic distance values increase rapidly ii) the probability of recovering true breed relationships is reduced. Among the breeds in this study the Sussex, Lincoln Red, White Park and Icelandic were distant from other breeds. They all have relatively small population sizes, low heterozygosities and a

small number of observed alleles, which indicates they are likely to have been subject to more genetic drift than other breeds. It is possible that a breed may appear genetically unique when, in fact, it perhaps carries only a subset of alleles encompassed by other related breeds.

Admixture or introgression also affects the relationships among breeds. Many European cattle breeds have been derived from crosses between older breeds (e.g. Guernsey, Blonde d'Aquitaine) and some populations have been subject to more recent introgression (e.g. Dairy Shorthorn). Using phylogenetic methods it is very difficult to separate the effect of admixture from that of common ancestry (Slatkin and Maddison 1990). However, the presence of admixed populations may not disturb the overall pattern of relationships among breeds. McDade (1992) concluded that in a phylogenetic analysis a hybrid will be placed closest to the parent with which it shares most characteristics, and that hybrids are unlikely to cause misrepresentation of the true relationships unless they are derived from very distantly related parents. In the analysis of 37 cattle breeds presented in this thesis the Murray Grey (Shorthorn crossed with Aberdeen Angus), for example, was found to be part of the British and North European group of breeds together with its parent breeds. However, it was not equidistant between its parent breeds and the unequal distance between them may have been due to the breed's small effective population size, which has resulted in substantial genetic drift.

Conclusions

It has been shown that European cattle breeds represent separate gene pools, and that although there may have been gene flow between breeds it has not been sufficient to prevent the breeds becoming genetically differentiated. Two major breed groups were identified, based on the blood type polymorphisms studied, with breeds from central and southern Europe (Mediterranean area) forming a separate group to breeds from northern Europe. In general, the relationships among breeds reflected their

geographic origin rather than the agricultural use for which the breeds have been selected.

Chapter 5

Genetic variation within the Hereford breed of cattle

5.1 Introduction

A primary objective of genetic conservation schemes is the maintenance of maximum genetic variation within species. In livestock species this can be achieved by the preservation of breeds (Hall and Bradley 1995). In recent years, changes in economic climate have promoted the use of breeds suited to intensive production systems, which has led to a few breeds becoming widespread while the breeds that they have replaced have declined in population size. In some cases native populations have been crossbred with imported stock in upgrading programmes. These native populations may have encompassed genetic variants that would have adaptive advantages under different environmental conditions. Introgression from other populations may result in the loss of original genetic variants, hence, breeds are not only threatened by extinction due to breed replacement but also by the genetic erosion of native populations. Several European breeds of cattle have been exported to other countries (e.g. USA, Canada, Australia), where successful breeding programmes have been developed. Over the past twenty years, European breeders have re-imported animals from other populations, in order to improve the performance of native populations. The Hereford, one of Britain's oldest native cattle breeds, is a good example of such a breed.

The Hereford originated in Herefordshire, England in the middle to late 1700s. Its ancestors consisted of cattle which were native to Britain at that time, but may also have included cattle from other regions of Europe (Heath-Agnew 1983). The Hereford breed has been under selection for more than 150 years. As one of the world's pre-eminent breeds it has proved successful in adapting to many environments and is now found in several countries throughout the world. In different countries the Hereford population will have been derived from separate founder groups and selection

objectives will have varied. These phenomena may have promoted genetic divergence of the different populations.

Modern Hereford populations can be divided by polled or horned status; the two groups generally being registered in separate herd book sections. Polled Herefords were created in the 1950s using crosses to the Red Poll in the USA and a Galloway bull in the UK (Heath-Agnew 1983). A survey of pedigree Hereford cattle born in Canada in 1982 (Koots and Crow 1989) found a clear subdivision of the population according to polled or horned status. Only 6% of matings were between parents of different horned-polled phenotype, or gave rise to a calf different in phenotype from its parents. Breeding strategies were also found to be different in the two groups; horned Hereford breeders used fewer imported sires (4%) than polled Hereford breeders (18% of sires were imported). This probably reflected greater use of artificial insemination in polled Herefords, for which 32,654 inseminations were recorded in Canada in 1987 compared with 237 for horned bulls (Koots and Crow 1989).

The level of genetic differentiation among populations will depend on the underlying population structures and the amount of migration between them. In a survey, based on herd book records of British and Irish horned Herefords, Ozkutuk and Bichard (1977) found considerable migration between herds, with 87% of all sires and 37% of all dams not homebred. However, 86% of sires were used in the country in which they had been born. Robertson and Asker (1951) had previously shown that the movement of cattle, particularly bulls, between herds was a directional process from a few "nucleus" herds to more numerous "multiplier" herds. Ozkutuk and Bichard (1977) confirmed the presence of a hierarchical structure among Hereford herds in Britain with, essentially, four herds at the nucleus level. One herd (Vern) was found to have made an overall genetic contribution of 24% to the breed. In Canada, Koots and Crow (1989) also found hierarchical structures in both polled and horned Hereford populations.

As more intensive style farming practices have become prevalent, continental breeds such as the Charolais or Limousin have overtaken the Hereford as the most numerous

beef breed in the UK. Estimated numbers of pure breeding Hereford females in 1969-70 were 12,700 for England, Wales and Scotland, and 9,200 for Ireland (Ozkutuk and Bichard 1977). In 1996 the number of breeding females in the UK was estimated at 4,000 (personal communication, Hereford Cattle Society). In response to demand for leaner, faster growing animals under intensive conditions, British Hereford breeders began to introduce Canadian bloodlines into their herds in the mid-1970s. This gene flow has been accelerated by the availability of artificial insemination and embryo transfer. The proportion of Herefords in the British population, born after 1980, that had one or more Canadian ancestors was estimated at 85-90% (personal observation from blood typing data).

Previous studies of herd book records have indicated that there are considerable levels of gene flow between herds, with a small number of "nucleus" herds providing the source of genes. This suggests that there should be little genetic differentiation between herds within national populations. The extent of genetic differentiation between national populations, however, will depend on the levels and pattern of gene flow between them. Gene flow may have occurred at a continuous low frequency (regular importation), or have taken place in a single burst (importation of a single group of founders). The longer time that populations have remained isolated the greater will be the genetic differentiation due to random genetic drift, particularly if the effective population size has been small. However, the significant level of importation of Canadian animals into the UK and other countries since the 1970s would suggest that Hereford populations may be becoming genetically more alike.

The objectives of this chapter were to use genetic markers (blood type polymorphisms) firstly, to assess the level of genetic differentiation between national populations of the Hereford and, secondly, to study the level of genetic differentiation between herds within the British population. The third objective was to examine whether there has been change in the genetic structure of the British Hereford population over time.

5.2 Materials and methods

5.2.1 Data

Blood type results for 3,448 horned and polled Herefords, collected between 1966 - 1996 by the Roslin Cattle Blood Typing Service were analysed in this study. The majority of animals had been blood typed for parentage verification (approximately 30%) or identification purposes (70%). An additional 254 blood samples were collected from ten British herds from animals born between 1985-1996. Samples from Irish, Canadian and New Zealand Herefords had been tested as part of the routine typing carried out by the Roslin Cattle Blood Typing Service. Blood typing data for Swedish Herefords were provided by the Blood Typing Laboratory at the Swedish University of Agricultural Sciences, Uppsala, Sweden.

The data were divided into 18 categories, according to animal's year of birth, ancestry, and horned or polled status. There were four categories of horned Hereford with 100% British ancestry, with animals born between 1960-69, 1970-79, 1980-89 and a modern day group regarded as representing the "traditional" British Hereford. Another group of horned animals was identified as being of "hybrid" ancestry (mixed British and Canadian). This group was further subdivided into three groups by year of birth, 1970-79, 1980-89, or 1990-96. Polled animals were simply classified into four groups according to their year of birth 1960-69, 1970-79, 1980-89, or 1990-96; their ancestry was predominantly British but also included some more recent Canadian ancestry. A large number of herds were represented in both the horned and polled groups. In the British horned group (both "hybrid" and 100% British ancestry) there were at least 270 herds represented (estimated from the number of herd prefixes) and in the polled group at least 210 herds were represented.

Animals with 100% Canadian ancestry were divided into four groups; horned animals born between 1960-79 or 1980-90, and polled animals born between 1960-79 or 1980-90. Animals from Ireland, New Zealand and Sweden were grouped by their country of

origin, although their ancestry will have included Herefords imported from other countries e.g. Britain and Canada. Data from six other breeds (Aberdeen Angus, Chianina, Limousin, Shorthorn, Simmental and Sussex) were included for comparison with the Hereford populations. The majority (> 80%) of animals included in the Irish, New Zealand and Swedish Hereford groups, and those from the six other breeds were born after 1980. Sample sizes for all the groups are shown in table 5.1.

Herds with a minimum of 25 blood typed animals were selected for the analysis of genetic differences between herds. Animals were defined as belonging to the herd in which they were bred, identified by the herd prefix. At any moment in time, a herd may be composed of animals from different sources, however, the gene flow between herds will be reflected in the next generation of animals bred within the herd. Data were available for 16 herds (6 horned and 10 polled), the number of animals sampled from each herd ranged between 25 and 122. At least 60% of animals sampled were born after 1980, and 75% of the remainder were born after 1970.

5.2.2 Allele frequencies

Allele frequencies were estimated at seven red cell antigen loci, A, B, C, F, L, S, Z, and the serum protein loci, transferrin and albumin. Full details of the antigens typed for each blood type system are given in chapter 2. An iterative procedure was used to estimate frequencies at dominant loci (Ceppellini et al, 1955; Weir, 1996), the procedure is described in more detail in chapter 4. At co-dominant loci (F system, transferrin and albumin) the allele frequencies were estimated by direct gene counting.

5.2.3 Heterozygosities and number of alleles

Expected heterozygosities (under Hardy Weinberg equilibrium) were calculated for each of the groups as $1-\sum p^2$ averaged over all loci. The observed number of alleles was expressed as a total number of alleles over all loci, since all loci (except transferrin) have only two alleles. Standard errors for the average heterozygosities and

number of alleles were obtained using bootstrapping, with 200 bootstrap replicates generated by resampling individual phenotypes.

5.2.4 Genetic distance

Genetic distances between national groups and between herds were calculated using the distance of Reynolds et al (1983). Standard errors of the genetic distances were obtained by bootstrapping (200 bootstrap replicates), resampling both loci and individual phenotypes (linked loci sampled together), see chapter 3 for details of the bootstrap technique. Multidimensional scaling (Manly 1986; Genstat 1993) was used to create graphical representations of the populations or herds from the distance matrices.

The Neighbour-joining method (Saitou and Nei 1987) was used to construct a dendrogram of population relationships from the genetic distance matrix using the program PHYLIP (Felsenstein 1995). The robustness of the dendrogram was evaluated using bootstrapping (100 replicates) by resampling loci, with linked loci sampled together.

5.2.5 Test of population differentiation

An exact test of differentiation (Hudson et al 1992) was performed on the different Hereford groups, both national populations and herds. Individuals from the groups to be compared were pooled and then randomly reassigned to groups without replacement, keeping the number of observations per group the same as in the original data. After each permutation of the data the genetic distance between groups was recalculated, and the number of times the new distance was greater than the observed distance counted to obtain the p values (probability of obtaining the observed genetic distance by chance). Ten thousand permutations of the data were carried out. A sequential Bonferroni test was applied, to adjust the 5% significance threshold to allow for multiple comparisons (Rice 1989). There were 276 pairwise comparisons between

national Hereford populations and other breeds, and 120 pairwise comparisons between herds.

5.2.6 Estimation of admixture in the "hybrid" population

Allele frequencies in crosses between populations can be predicted according to the proportion of each parental breed in the crossbreed population (Long 1991). A weighted least squares estimate of admixture can be obtained by re-writing the allele frequency prediction equation. If P_{hi}, P_{1i} and P_{2i} represent the frequency of allele i in the hybrid population, the first parental population and the second parental population respectively, and ε_i is the error due to sampling and genetic drift then:

$$(P_{hi} - P_{2i}) = \mu(P_{1i} - P_{2i}) + \varepsilon_i \qquad [6.10]$$

where μ is the proportional contribution of the first parental population.

Long and Smouse (1983) and Long (1991) proposed an iterative procedure for obtaining the weighted least squares estimate. Chakraborty et al (1992) presented closed form expressions for the estimate of admixture and its standard error, which were utilised here. The estimate of admixture is given by:

$$\hat{\mu} = \frac{\sum_{l=1}^{L}\sum_{i=1}^{r_l}(P_{P1i}^{(l)} - P_{P2i}^{(l)})(P_{hi}^{(l)} - P_{P2i}^{(l)})/P_{hi}^{(l)}}{\sum_{l=1}^{L}\sum_{i=1}^{r_l}(P_{P1i}^{(l)} - P_{P2i}^{(l)})^2/P_{hi}^{(l)}} \qquad [6.11]$$

where L is the number of total number of loci, and r_l is the number of alleles at locus l. The sampling variance of $\hat{\mu}$ is:

$$V(\hat{\mu}) = \frac{MSE}{\sum_{l=1}^{L}\sum_{i=1}^{r_l}(P_{P1i}^{(l)} - P_{P2i}^{(l)})^2/P_{hi}^{(l)}} \qquad [6.12]$$

where MSE (mean square error) is:

$$\frac{\sum_{l=1}^{L}\sum_{i=1}^{r_l}[(P_{hi}^{(l)} - P_{P2i}^{(l)}) - (\hat{\mu}P_{P1i}^{(l)} - P_{P2i}^{(l)})]^2 / P_{hi}^{(l)}}{r - L}$$

and $r = \sum_{l=1}^{L} r_l$ is the total number of alleles at all L loci.

An estimate of the proportion of Canadian genes in the "hybrid" (British cross Canadian) Herefords that were born between 1980-96, and the standard error of the admixture estimate were obtained using equations 6.11 and 6.12. Parental populations contributing to the "hybrid" population were taken as the British 1960s group and the Canadian Herefords (1960s and 1980s grouped together).

5.3 Results

5.3.1 Heterozygosity and number of alleles

Allele frequencies for each of the Hereford populations are shown in Appendix II. Expected average heterozygosities and number of alleles observed in each population, together with their standard errors are shown in table 5.1. The expected average heterozygosities ranged between 0.19 (± 0.062) and 0.26 (± 0.074). Lowest heterozygosities were observed in the Canadian polled and in the Canadian horned groups (1960s). Among the British groups the least heterozygous was the "traditional" Hereford. Highest heterozygosities were observed in the "hybrid" animals (1970s), and in Swedish Herefords. Other breeds in the study generally had higher heterozygosities than the Hereford groups, with the exception of the Sussex. Figure 5.1 shows the change over time in the average heterozygosity of the British and Canadian Hereford populations.

Table 5.1 Expected average heterozygosities, and observed number of alleles with their respective standard errors (in parentheses) for 18 Hereford populations and 6 other breeds.

	No. animals	Average heterozygosity	Observed no. alleles
Hereford populations			
British 1960-69 (B60)	388	0.24 (0.071)	57 (0.58)
British 1970-79 (B70)	271	0.25 (0.071)	53 (1.00)
British 1980-89 (B80)	26	0.22 (0.068)	49 (0.73)
Traditional (TR)	61	0.21 (0.073)	48 (0.63)
Hybrid 1970-79 (H70)	46	0.25 (0.065)	53 (1.53)
Hybrid 1980-89 (H80)	166	0.23 (0.064)	56 (0.83)
Hybrid 1990-96 (H90)	57	0.23 (0.041)	51 (0.57)
Poll 1960-69 (P60)	125	0.24 (0.072)	57 (0.61)
Poll 1970-79 (P70)	425	0.23 (0.070)	58 (0.95)
Poll 1980-89 (P80)	381	0.22 (0.068)	58 (0.63)
Poll 1990-96 (P90)	117	0.22 (0.057)	56 (0.51)
Canadian 1960-79 (CN60)	25	0.20 (0.048)	53 (1.44)
Canadian 1980-96 (CN80)	65	0.21 (0.058)	56 (1.08)
Canadian Poll 1960-79 (CP60)	26	0.19 (0.062)	49 (1.03)
Canadian Poll 1980-96 (CP80)	39	0.19 (0.050)	51 (0.81)
Irish (IR)	565	0.23 (0.061)	61 (0.77)
Swedish (SW)	527	0.26 (0.074)	60 (0.81)
New Zealand (NZ)	158	0.23 (0.067)	59 (0.85)
Other breeds			
Aberdeen Angus (AA)	197	0.24 (0.046)	62 (0.69)
Chianina (CI)	66	0.29 (0.057)	64 (0.00)
Limousin (LM)	1813	0.32 (0.025)	64 (0.00)
Shorthorn (SH)	64	0.27 (0.046)	59 (0.79)
Simmental (SM)	928	0.24 (0.049)	64 (0.00)
Sussex (SX)	64	0.22 (0.040)	61 (0.86)

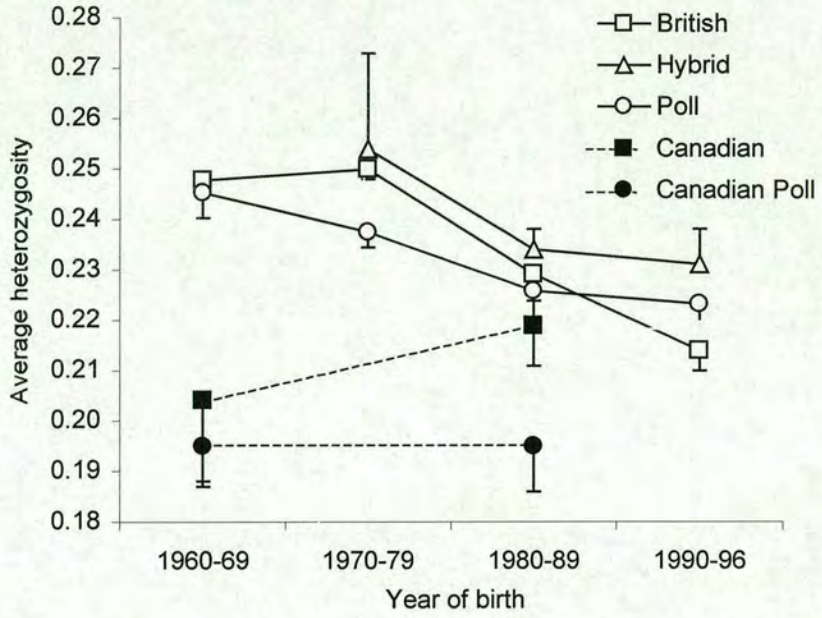

Figure 5.1. Change over time in the average heterozygosity of the British and Canadian Hereford populations

The Canadian Hereford groups had lower heterozygosities than the British groups, with their heterozygosities not seeming to change much between 1960 and 1989. The apparent slight rise in heterozygosity in the Canadian Herefords may be due to sampling. In contrast, the heterozygosity of the British Hereford groups has declined between 1960 and 1996. Crossing of British and Canadian animals in the 1970s resulted in an increase of heterozygosity in the "hybrid" population, but the level of heterozygosity in the "hybrid" population is now falling. The "hybrid" population is, nevertheless, currently more heterozygous than the group with pure British ancestry ("traditional"). This is due to the severe reduction in numbers of animals of pure British ancestry, from several thousand in the 1960s to around 400 in the 1990s. In addition, the modern population of British animals consists of a limited number of families, compared with the original 1960s population. The population of poll Herefords in the UK has a heterozygosity that is currently mid-way between that of the "hybrid" and pure British Hereford populations. Number of alleles observed ranged from 48 in the " traditional" group to 60 in Swedish, and 61 in Irish Herefords.

5.3.2 Genetic distances

Genetic distances between Hereford populations and the other breeds are shown in table 5.2. Genetic distances between the Hereford populations ranged between 0.011 (± 0.009) and 0.102 (± 0.052). Distances between the Hereford populations and other breeds were greater overall, ranging from 0.067 (± 0.038) between the Shorthorn and Poll Herefords (1970s) to 0.289 (± 0.057) between the Chianina and Canadian Herefords (1960s) or Hybrid Herefords (1990s). Multidimensional scaling (figure 5.2) indicated that the Hereford populations group together and are distinct from other breeds. Figure 5.3 shows the decline in the stress statistic as the number of dimensions of the solution is increased. The solution presented is a three-dimensional solution, which has a stress statistic of just below 0.06. This value is considered to be reasonably low (Manly 1986), indicating that the three-dimensional solution provides a good representation of the population relationships. There was no clear grouping of horned and polled populations. There was, however, a pattern associated with the geographic origin of the Hereford groups. The British populations grouped together, irrespective of their ancestry and polled or horned status, similarly the Canadian groups formed a cluster. The Irish, New Zealand and Swedish populations were peripheral to the British populations, with the Swedish group separated from the Irish and New Zealand groups.

The neighbour joining tree (figure 5.4) also showed the Hereford populations formed a cluster which is distinct from other breeds. In contrast with the multidimensional scaling there is no clear distinction between the Hereford groups of different geographic origin, for example, the Canadian groups are clustered with four British groups and the Swedish group. However, bootstrap values in this portion of the tree are generally low (8-58%) indicating that these groupings are not robust, and particularly not when compared with the division of the Hereford groups from other breeds (bootstrap values ranging between 25-92%).

Table 5.2 Genetic distances (below diagonal) and standard errors (above diagonal) for 18 Hereford populations and 6 other breeds. Shaded values were found to be significantly different from zero ($p < 0.05$) using a permutation test.

	No animals	TR	B60	B70	B80	H70	H80	H90	P60	P70	P80	P90
TR	61		0.014	0.018	0.027	0.016	0.016	0.034	0.017	0.015	0.015	0.013
B60	388	0.022		0.013	0.026	0.023	0.020	0.022	0.010	0.021	0.020	0.016
B70	271	0.029	0.023		0.030	0.009	0.015	0.031	0.018	0.005	0.010	0.028
B80	26	0.032	0.029	0.035		0.035	0.039	0.039	0.031	0.036	0.025	0.025
H70	46	0.025	0.035	0.011	0.037		0.017	0.036	0.019	0.014	0.016	0.027
H80	166	0.029	0.028	0.025	0.041	0.020		0.025	0.008	0.016	0.007	0.015
H90	57	0.080	0.072	0.094	0.083	0.088	0.054		0.026	0.034	0.028	0.029
P60	125	0.029	0.016	0.031	0.032	0.030	0.013	0.055		0.021	0.009	0.012
P70	425	0.035	0.039	0.011	0.044	0.015	0.022	0.088	0.029		0.010	0.027
P80	381	0.028	0.033	0.029	0.031	0.023	0.016	0.066	0.017	0.015		0.010
P90	117	0.027	0.034	0.053	0.034	0.049	0.031	0.047	0.020	0.044	0.016	
CN60	25	0.064	0.075	0.063	0.066	0.046	0.036	0.058	0.044	0.039	0.022	0.035
CN80	65	0.038	0.052	0.053	0.054	0.041	0.026	0.043	0.032	0.038	0.019	0.018
CP60	26	0.055	0.068	0.063	0.038	0.061	0.069	0.094	0.060	0.048	0.028	0.034
CP80	39	0.054	0.080	0.073	0.052	0.067	0.082	0.105	0.077	0.058	0.042	0.044
IR	565	0.040	0.044	0.038	0.063	0.033	0.020	0.048	0.028	0.032	0.028	0.030
SW	527	0.080	0.069	0.088	0.082	0.085	0.062	0.083	0.060	0.082	0.066	0.066
NZ	158	0.061	0.068	0.047	0.081	0.043	0.035	0.105	0.039	0.024	0.037	0.057
AA	197	0.151	0.156	0.109	0.156	0.111	0.110	0.164	0.130	0.090	0.115	0.142
CI	66	0.273	0.254	0.211	0.259	0.244	0.250	0.289	0.257	0.223	0.259	0.273
LM	1813	0.187	0.181	0.137	0.192	0.147	0.154	0.196	0.178	0.133	0.168	0.201
SH	64	0.117	0.127	0.082	0.112	0.077	0.082	0.133	0.104	0.067	0.088	0.118
SM	928	0.177	0.162	0.124	0.197	0.145	0.135	0.182	0.155	0.124	0.155	0.179
SX	64	0.148	0.141	0.121	0.166	0.120	0.095	0.173	0.117	0.107	0.115	0.152

Table 5.2, continued. Genetic distances (below diagonal) and standard errors (above diagonal) for 18 Hereford populations and 6 other breeds. Shaded values were found to be significantly different from zero ($p < 0.05$) using a permutation test.

	No animals	CN60	CN80	CP60	CP80	IR	SW	NZ	AA	CI	LM	SH	SM	SX
TR	61	0.043	0.019	0.030	0.032	0.017	0.047	0.031	0.042	0.083	0.035	0.041	0.044	0.058
B60	388	0.053	0.025	0.037	0.037	0.015	0.044	0.033	0.045	0.062	0.034	0.042	0.040	0.057
B70	271	0.036	0.024	0.030	0.034	0.022	0.043	0.022	0.026	0.063	0.031	0.030	0.023	0.037
B80	26	0.043	0.027	0.027	0.036	0.032	0.047	0.062	0.069	0.084	0.056	0.058	0.060	0.091
H70	46	0.026	0.023	0.038	0.043	0.024	0.045	0.021	0.045	0.073	0.044	0.042	0.044	0.039
H80	166	0.029	0.013	0.044	0.045	0.010	0.045	0.020	0.044	0.052	0.032	0.030	0.036	0.028
H90	57	0.034	0.025	0.042	0.041	0.023	0.047	0.047	0.064	0.047	0.071	0.044	0.080	0.070
P60	125	0.035	0.015	0.039	0.039	0.010	0.045	0.024	0.050	0.064	0.037	0.038	0.042	0.036
P70	425	0.029	0.020	0.034	0.033	0.019	0.044	0.015	0.032	0.068	0.029	0.038	0.022	0.029
P80	381	0.020	0.011	0.028	0.027	0.013	0.044	0.019	0.044	0.070	0.027	0.043	0.029	0.039
P90	117	0.030	0.011	0.024	0.027	0.015	0.048	0.036	0.047	0.061	0.035	0.031	0.044	0.056
CN60	25		0.019	0.041	0.044	0.027	0.051	0.048	0.051	0.057	0.054	0.044	0.062	0.063
CN80	65	0.017		0.028	0.029	0.009	0.050	0.028	0.036	0.058	0.033	0.029	0.041	0.051
CP60	26	0.040	0.038		0.016	0.035	0.048	0.069	0.060	0.083	0.052	0.058	0.060	0.095
CP80	39	0.053	0.045	0.012		0.036	0.052	0.067	0.044	0.076	0.047	0.047	0.060	0.093
IR	565	0.038	0.016	0.067	0.079		0.045	0.019	0.024	0.046	0.022	0.021	0.028	0.040
SW	527	0.083	0.063	0.098	0.102	0.076		0.045	0.030	0.064	0.020	0.033	0.030	0.028
NZ	158	0.058	0.048	0.085	0.094	0.038	0.029		0.052	0.055	0.039	0.040	0.052	0.044
AA	197	0.116	0.102	0.146	0.149	0.086	0.156	0.083		0.046	0.016	0.025	0.017	0.024
CI	66	0.289	0.261	0.280	0.281	0.227	0.263	0.226	0.031		0.041	0.051	0.049	0.056
LM	1813	0.185	0.166	0.205	0.209	0.144	0.172	0.143	0.012	0.129		0.024	0.016	0.045
SH	64	0.097	0.085	0.118	0.121	0.076	0.132	0.074	0.016	0.170	0.016		0.030	0.044
SM	928	0.173	0.142	0.205	0.219	0.100	0.186	0.125	0.011	0.152	0.014	0.112		0.054
SX	64	0.150	0.142	0.199	0.203	0.126	0.142	0.107	0.050	0.262	0.035	0.118	0.176	

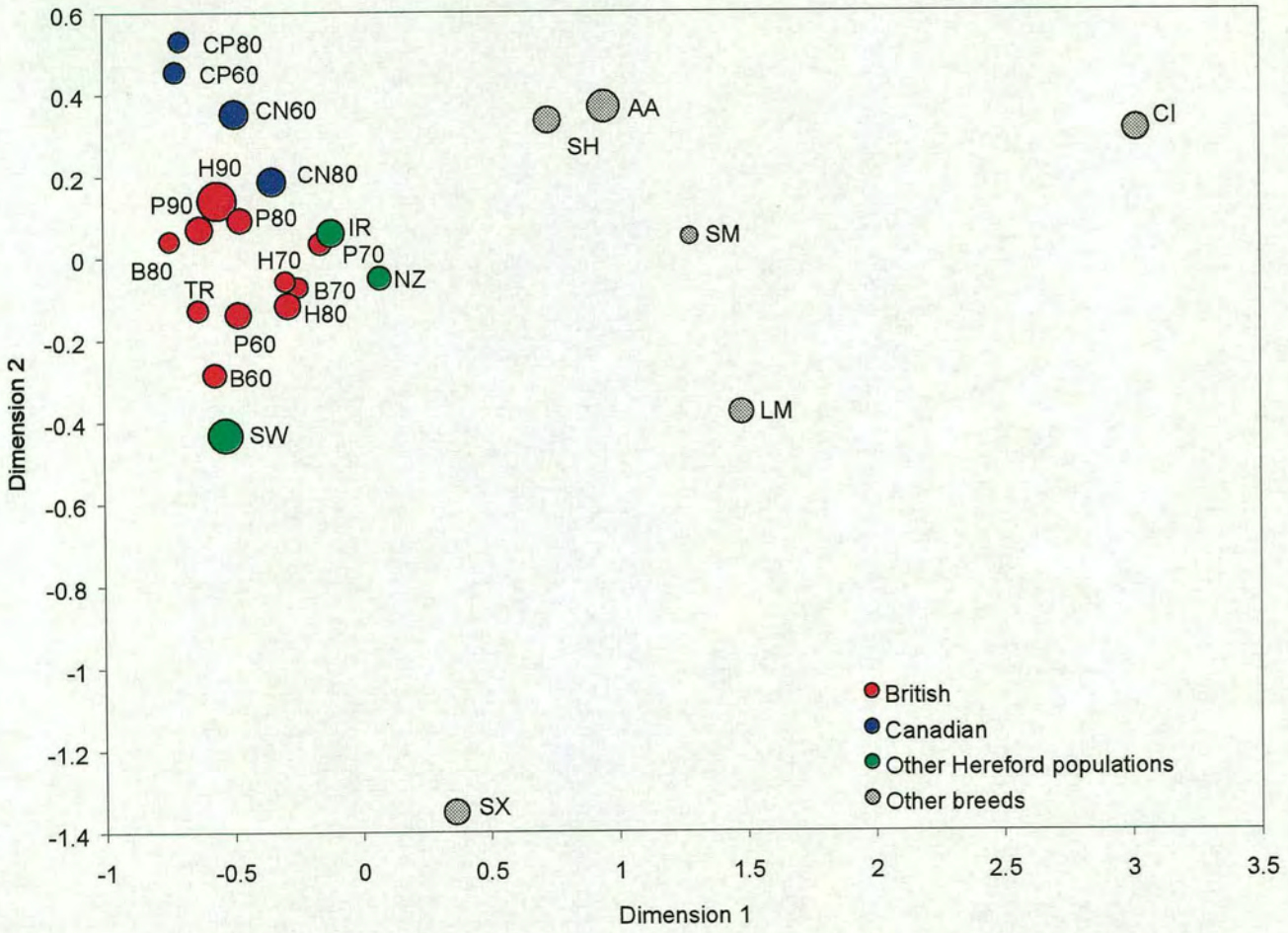

Figure 5.2. Three-dimensional representation of the genetic distances among 18 Hereford populations and 6 other breeds, obtained using multidimensional scaling. Distance in the third dimension is represented by the diameter of the point, so points that are distant from the viewer are smaller than those that are closer.

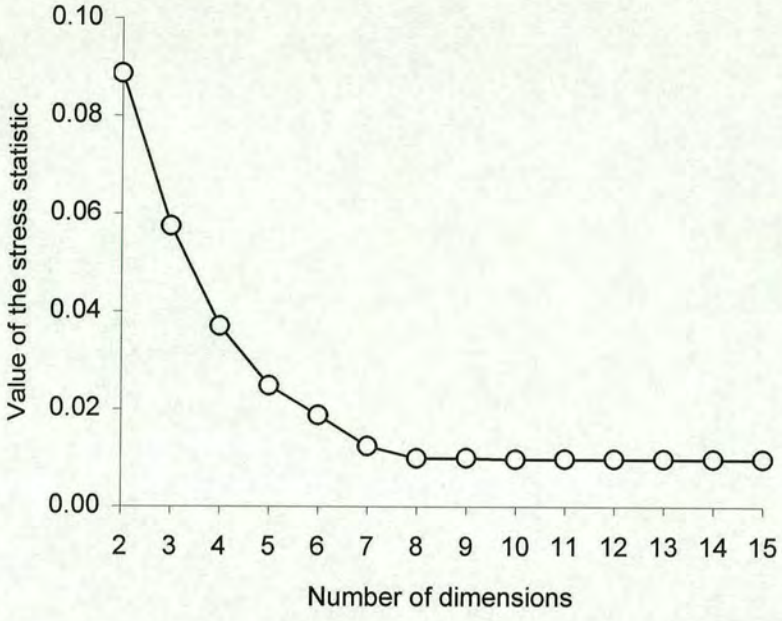

Figure 5.3. Decline in the stress statistic as the number of dimensions of the multidimensional scaling solution is increased

5.3.3 Differentiation between national populations

All the Hereford populations were significantly different from the six other breeds tested (Aberdeen Angus, Chianina, Limousin, Shorthorn, Simmental, and Sussex), $p < 0.00001$. The Irish, New Zealand and Swedish populations were also significantly different from most other Hereford groups ($p < 0.01$), although the Irish and New Zealand Herefords were not significantly different from some polled groups. In general, the Canadian populations were significantly different from populations with 100% British ancestry, but were not significantly different from the "hybrid" or poll populations. The British 1960-69 group was significantly different from the "hybrid" populations and not significantly different from groups with 100% British ancestry. However, the more recent British groups (1970-79, 1980-89 and "traditional") were not significantly different from the "hybrid" groups, and some were not significantly different from the poll populations.

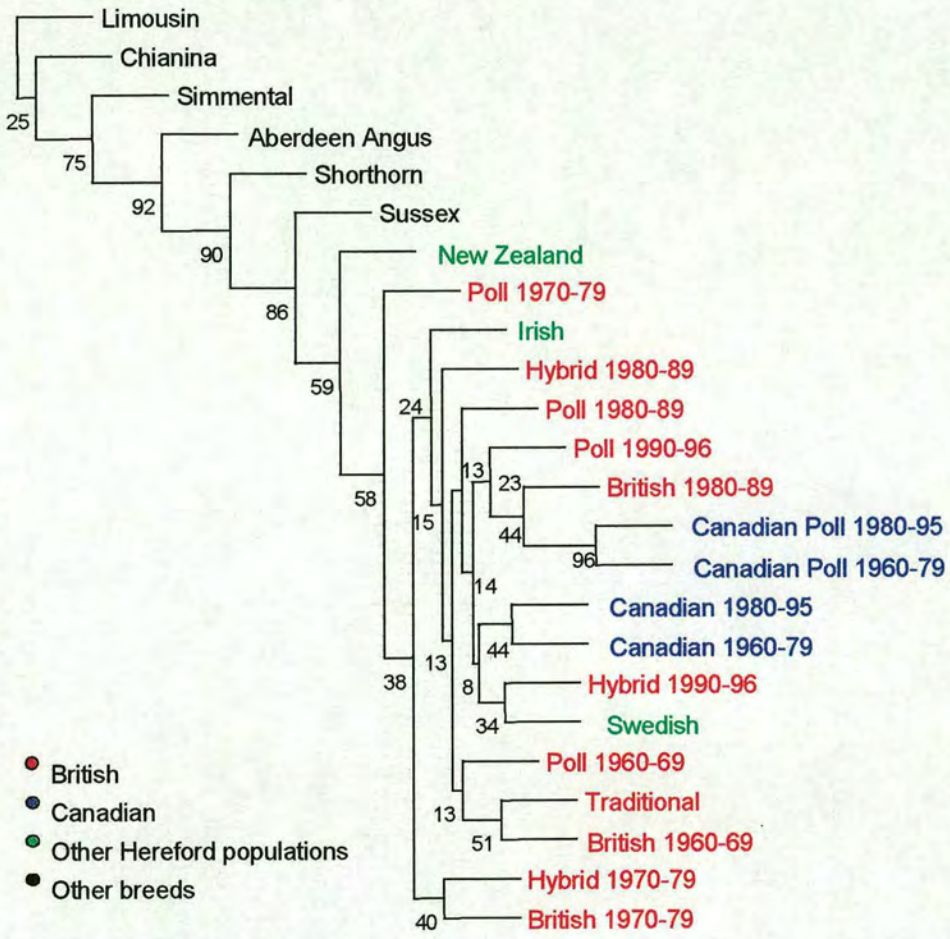

Figure 5.4 Neighbour joining tree showing the relationships among 18 Hereford populations and 6 other breeds. Bootstrap proportions (number of times a node was observed in 100 replicates) are shown below each node.

5.3.4 Differentiation between herds

The genetic distances between herds are shown in table 5.3, and the representation of the distance matrix obtained using multidimensional scaling is shown in figure 5.5. The Canadian and British (1960s) populations are included in the multidimensional scaling diagram, as points of reference. There was no clear grouping of polled and horned herds. The majority of the herds formed part of a central grouping, with only five outlying herds. The permutation test indicated that most herds were significantly different from one another.

Table 5.3 Genetic distances (below diagonal) and standard errors (above diagonal) for 16 Hereford herds (6 horned (H) and 10 polled (P)). Shaded values were found to be significantly different from zero ($p < 0.05$) using a permutation test.

	No animals	H1	H2	H3	H4	H5	H6	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
H1	49		0.043	0.051	0.031	0.047	0.048	0.033	0.053	0.052	0.022	0.064	0.047	0.044	0.051	0.031	0.045
H2	73	0.068		0.034	0.018	0.036	0.048	0.017	0.025	0.033	0.030	0.019	0.111	0.058	0.052	0.037	0.024
H3	51	0.082	0.061		0.020	0.026	0.043	0.034	0.024	0.040	0.041	0.045	0.087	0.050	0.060	0.026	0.027
H4	100	0.049	0.039	0.046		0.023	0.047	0.027	0.020	0.024	0.025	0.032	0.074	0.032	0.035	0.028	0.022
H5	41	0.083	0.057	0.033	0.043		0.059	0.019	0.033	0.050	0.027	0.046	0.068	0.041	0.044	0.020	0.028
H6	36	0.123	0.116	0.142	0.093	0.138		0.053	0.059	0.045	0.059	0.054	0.071	0.045	0.074	0.069	0.056
P1	122	0.067	0.034	0.053	0.044	0.039	0.105		0.019	0.032	0.016	0.029	0.081	0.027	0.040	0.025	0.028
P2	44	0.084	0.041	0.033	0.041	0.054	0.137	0.042		0.045	0.034	0.036	0.126	0.063	0.057	0.025	0.026
P3	47	0.112	0.062	0.089	0.059	0.110	0.136	0.075	0.073		0.038	0.033	0.115	0.055	0.056	0.045	0.035
P4	79	0.051	0.064	0.050	0.061	0.054	0.134	0.025	0.040	0.093		0.045	0.065	0.047	0.039	0.021	0.033
P5	36	0.101	0.046	0.082	0.055	0.088	0.172	0.065	0.059	0.064	0.084		0.127	0.075	0.064	0.043	0.033
P6	29	0.073	0.131	0.112	0.090	0.072	0.152	0.066	0.134	0.175	0.064	0.157		0.067	0.083	0.076	0.100
P7	36	0.082	0.091	0.124	0.067	0.111	0.090	0.057	0.112	0.116	0.079	0.122	0.084		0.038	0.052	0.048
P8	29	0.089	0.089	0.086	0.059	0.058	0.108	0.033	0.065	0.120	0.043	0.100	0.072	0.056		0.043	0.052
P9	28	0.060	0.061	0.032	0.033	0.036	0.129	0.034	0.024	0.079	0.030	0.078	0.083	0.087	0.046		0.027
P10	25	0.071	0.038	0.032	0.028	0.042	0.130	0.038	0.034	0.063	0.056	0.046	0.104	0.083	0.064	0.034	

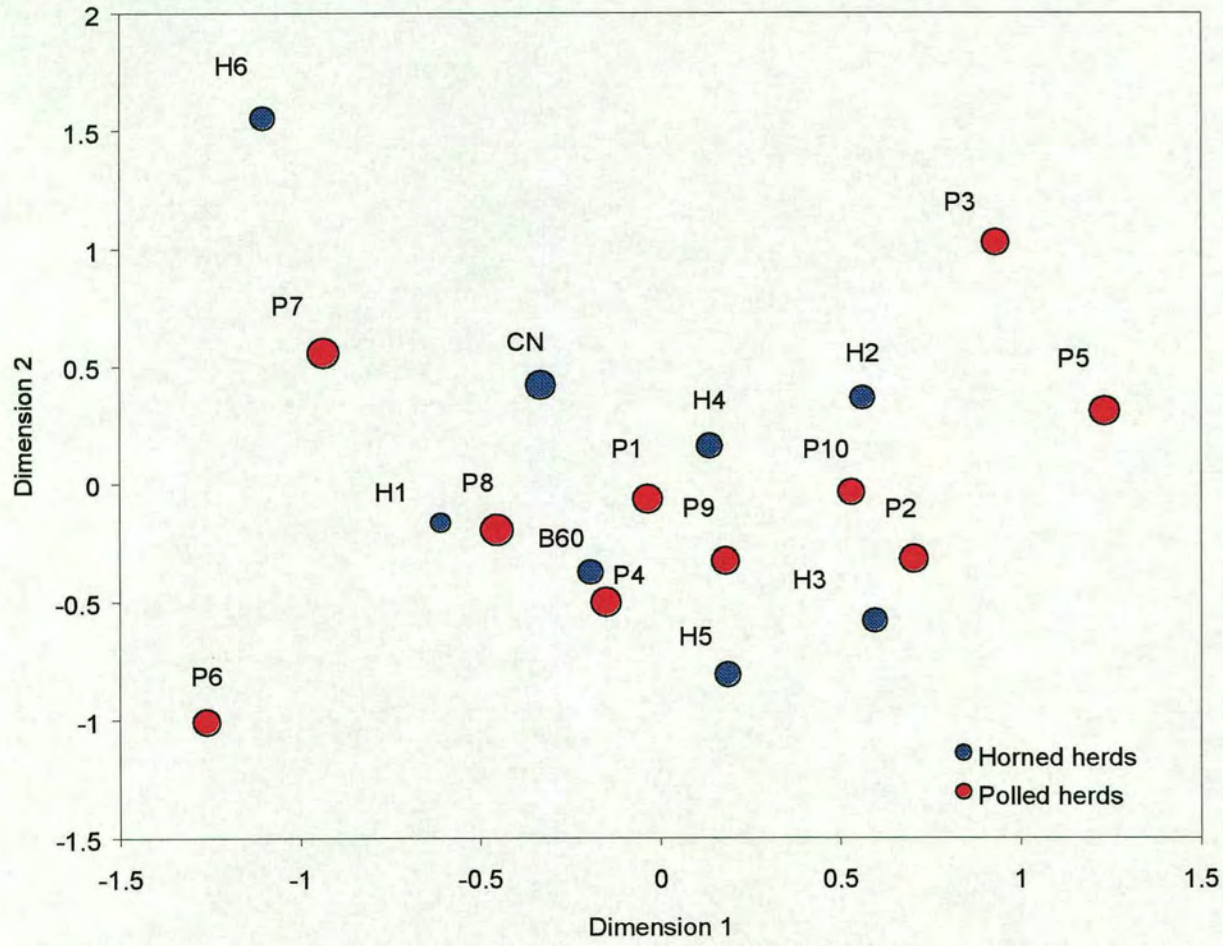

Figure 5.5. Three-dimensional representation of the genetic distances among 6 horned Hereford herds (designated by an H) and 10 polled herds (designated by a P), obtained using multidimensional scaling. The Canadian (CN) and British 1960s (B60) are shown to indicate the relative positions of the parental populations that contributed to the modern "hybrid" population in Britain. Distance in the third dimension is represented by the diameter of the point, so points that are distant from the viewer are smaller than those that are closer.

5.3.5 Estimate of admixture in the "hybrid" population

The proportion of Canadian genes in the "hybrid" (1980-90s) population was estimated at 0.65 ($\pm$ 0.21) using the point estimator of Chakraborty et al (1992). Figure 5.6 shows the difference in allele frequencies between the British (1960s) population and the "hybrid" (1980-90s) group plotted against the difference in allele frequencies between the two parental populations (British (1960s) and Canadian). The slope of the line fitted to the points gives an estimate of the proportion of genes from the Canadian parental population (Chakraborty et al 1992). A regression coefficient of 0.59 was obtained, which is in relatively close accordance with the point estimate of admixture. As the parental populations are closely related (share very recent common ancestry) the difference between their allele frequencies is small, which results in a larger standard error for the estimate of admixture than if the parental populations had been more distantly related.

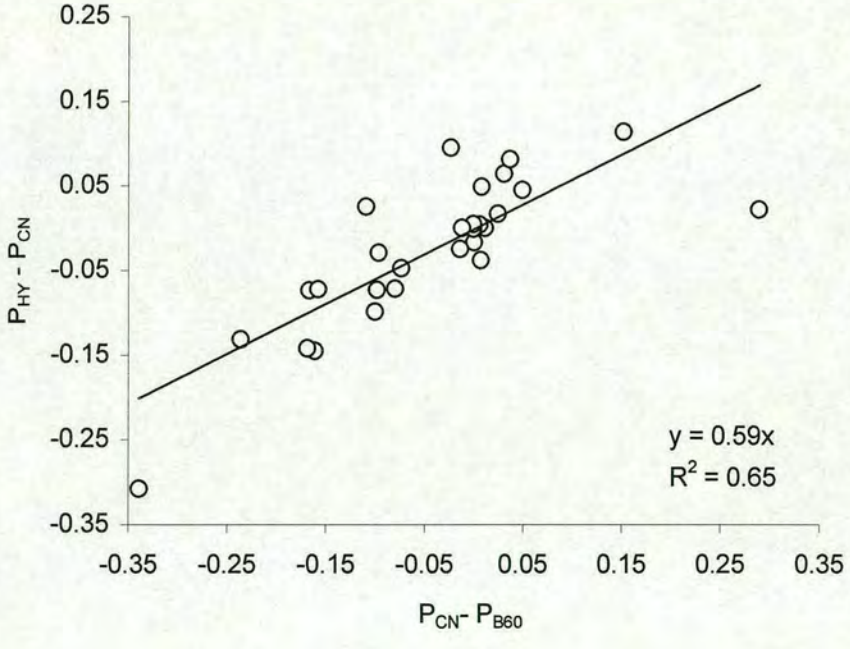

Figure 5.6. Relationship between P_{CN} - P_{B60} and P_{HY} - P_{CN} (where P_{CN} and P_{B60} are the allele frequencies in the Canadian and British Hereford parental populations and P_{HY} is the allele frequency in the "hybrid" Hereford population) for 7 blood type systems (27 red cell antigens) and 2 serum protein loci. The fitted line has a slope equal to $\hat{\mu}$=0.59 which is the estimate of admixture.

5.4 Discussion

There is significant genetic differentiation between Hereford populations from different geographical regions. However, when compared with other breeds the Hereford populations clustered together, i.e. the differences between Hereford groups are not as large as between breed differences. Similarly, differences between the British polled and horned populations were not as large as the differences between Herefords from different countries.

There were also significant differences between Hereford herds within the British population. Without pedigree information it is not easy to say why herds are different. However, there was variation in the breeding strategies used within herds, some herds consisting only of animals with 100% British ancestry, some breeders favouring the use of Canadian animals to a greater or lesser extent. At least one of the herds was known to use as sires, almost exclusively, animals with Canadian ancestry which were not related to other sires used in other herds. The significant differences could also arise as a result of the failure of the permutation assumptions, i.e. that animals are independent, since animals within herds tend to be related. In other words there are significant differences between herds because there is a family structure that is generally confounded with herd.

Canadian Herefords were found to be less heterozygous than other Hereford groups. This may be as a result of founder effect, or perhaps due to higher levels of selection (and a related increase in the level of inbreeding). The higher heterozygosities observed in the "hybrid" population of the 1970s, and in the Swedish population are perhaps due to the introduction of Canadian animals into breeding programmes. The crossing of two previously isolated populations would result in an increase of heterozygosity. However, the introduction of Canadian Herefords into other national populations, while resulting in a short-term reduction in inbreeding (increase in heterozygosity), could in the longer term cause a greater loss of genetic variation if

complete substitution occurs. At present, the level of Canadian Hereford genes in the British population is estimated at about 65%.

Founder effects and drift may have played a significant part in the differentiation of national populations. In addition to these effects, intense selection may have further reduced levels of genetic variation in some populations. Studies on the effect of introgression of Holstein genes into European populations of black-and-white cattle have found that while the performance of production traits (milk yield and protein) was improved, there was an unfavourable effect on fertility traits (Lidauer and Mantysaari 1996). If the long-term effect of admixture is to cause a reduction in genetic variation, then the complete replacement of the British Hereford population with Canadian animals would be disadvantageous. Further studies are required to look at the long-term effects of crossing national populations on genetic variation and performance, in order to assess whether the short-term benefits outweigh the long-term costs. Conservation of genetic variation in British Hereford populations should be considered by breeders, in the interests of the long-term future of the breed.

Chapter 6

Discriminating among cattle breeds using genetic markers

6.1 Introduction

Differences in the evolutionary histories of livestock breeds have resulted in the formation of genetically distinct populations. Genetic markers have been shown to be useful tools for assessing breed diversity and relationships (Baker and Manwell 1991; MacHugh et al 1994), and could provide a potentially powerful method of distinguishing between individuals of different breeds. DNA markers, such as microsatellites, have been found to be particularly useful for discriminating among populations. For example, individuals from fourteen human populations were found to cluster together in a phylogenetic tree according to their continent of origin, with 87.8% of individuals clustering with other individuals from the same population (Bowcock et al 1994). This observation was based on thirty microsatellite markers typed in about ten individuals per population. MacHugh (1996) similarly observed that individual cattle from the same breeds tended to cluster together, based on their genotypes at twenty microsatellite loci. Buchanan et al (1994) found that eight microsatellite loci were sufficient to distinguish between six breeds of sheep with between 88% and 99% accuracy, depending on the breed. Individuals were allocated to the breed that was most probable, based on the allele frequencies observed in the breeds. These results suggest that genetic markers could be used to identify the breed of an individual animal.

A genetic test of breed identity would be valuable for several reasons. Livestock genetic conservation is currently based on the maintenance of breeds and a fundamental principle of breed conservation is the promotion of pure-breeding, which has traditionally been achieved by keeping pedigree records (Hall and Bradley

1995). However, crossbreeding or the introgression of genes from other breeds is a common strategy in livestock breeding programmes. Genetic markers could provide a means of estimating the proportion of a genome that is characteristic of the parental breeds, and consequently a method of identifying when introgression has taken place (Bradley et al 1994). This application may be useful for checking whether or not populations are pure breeding, which is an important consideration in the genetic characterization of livestock breeds. A test of breed identity based on genetic markers would also be valuable for the validation of livestock products. This application may become increasingly important as breed names become more widely used as a "brand" name for livestock products. Beef and milk from several cattle breeds are already being marketed under the breed name (e.g. Aberdeen Angus, Highland, Welsh Black beef, Ayrshire, Jersey and Guernsey milk, and several rare breeds within the Rare Breeds Meat Marketing Scheme). Protection of the brand name may require that products sold under it can be validated by a DNA test.

Different genetic markers will vary in their informativeness for distinguishing between breeds. For example, allozyme loci are generally less variable and hence less informative than microsatellite or minisatellite markers. Identification based on electrophoretic analysis of protein variation has been used to estimate stock compositions in various species of salmon (Shaklee and Varnavskaya 1994; Van Doornik et al 1996). While allozyme loci can be used to distinguish between distantly related populations (Davidson et al 1989), highly variable DNA markers, such as minisatellites, were found to be more powerful discriminators for closely related salmon populations (Galvin et al 1995; Miller et al 1996). However, Bowcock et al (1994) noted that loci with higher values of F_{ST}, which they found to be negatively correlated with locus heterozygosity, tended to produce phylogenetic trees with closer clustering of individuals from the same geographic origin. They consequently suggested that loci with a high diversity (high heterozygosity and large number of alleles), might be less informative than loci with a lower diversity.

New technologies that allow large numbers of samples to be easily genotyped, and that have been developed for forensic application are now becoming available. One such technique is the typing of single nucleotide polymorphisms (SNPs) using a semi-automated assay (Nikiforov et al 1994; Delahunty et al 1996). These polymorphisms are usually diallelic (Delahunty et al 1996), and although this means that they are inherently less informative than markers with multiple alleles, they have several advantages for forensic testing. These advantages being that they are widely distributed throughout the genome (Kwok et al 1994), can be more reliably amplified using PCR than repeat variations (such as microsatellites) and are easier to automate on a large scale (Nickerson et al 1990; Nikiforov et al 1994). Since SNPs are diallelic and co-dominant, allele frequencies are easy to determine and can be estimated in a population in a number of ways (Syvanen et al 1992; Kwok et al 1994). Allele frequencies at nineteen SNPs, determined from 76 individuals from the reference pedigrees for the Human Genome Mapping Project, ranged between 0.11 and 0.89 (Delahunty et al 1996). Single nucleotide polymorphisms also occur in livestock species, however no estimates of allele frequencies in cattle breeds have yet been published.

The objective of this chapter was to compare two types of marker, diallelic polymorphisms and microsatellites, for their efficiency in distinguishing among cattle breeds and to investigate the number of markers required to distinguish between purebred European breeds and their crosses. Published estimates of microsatellite allele frequencies in various cattle breeds are available. Since no published estimates of allele frequencies for SNPs in cattle were available, the blood type data (described in chapters 2 and 4) were used to represent the distribution of allele frequencies at a set of diallelic markers for the different breeds. Markers were compared by using discriminant analysis to formulate an allocation rule. Allocation of individuals to breeds was based on genotype probabilities, derived from the marker allele frequencies in the different breeds. Individuals were allocated to the breed for which the probability was highest. This type of allocation rule is one method of discriminant analysis.

Other methods of discriminating between populations, based on the distances between individuals of the same population and individuals from different populations, have also been derived. Bowcock et al (1994) and MacHugh (1996) looked at how individuals from the same population clustered within a phylogenetic tree. The tree was based on the number of shared alleles between individuals. More formal methods of discriminant analysis based on distances have been described by Cuadras (1989, 1992). The approach to discriminant analysis utilising genotype probabilities is a parametric method based on the distribution of gene frequencies. The approach based on genetic distances between individuals is non-parametric, i.e. based on observed distances, which do not have a known distribution. In the analysis described in this chapter a parametric approach was taken as, in general, parametric methods are more powerful than non-parametric methods. The possible disadvantage of using a parametric approach is that it is dependent on a model, for example, the estimation of a breed probability from gene frequencies assumes random mating, Hardy-Weinberg genotype frequencies (although with co-dominant markers the observed genotype frequencies could be used), and independence of markers. A non-parametric method would not be subject to these assumptions, and in situations where the assumptions are likely to be broken would perhaps be a more flexible approach.

The two types of marker, microsatellite and diallelic markers, were compared using two data sets that had seven European breeds in common. Power of discrimination, for each marker set, was assessed via the error rates (probability of misclassification). The number of each type of marker required to discriminate among breeds for a given error rate was also investigated. If a commercial test of breed is to be developed, it may also be desirable to be able to identify crossbred animals. A number of breed meat marketing schemes (e.g. Aberdeen Angus marketing in the UK, and the "Hereford Prime" scheme in New Zealand) not only accept pure-bred animals, but 50% crosses. The number of markers required to identify crossbred

animals was investigated by simulating microsatellite data for crosses between the Friesian and four other breeds; Aberdeen Angus, Charolais, Hereford and Simmental.

6.2 Materials and methods

6.2.1 Theoretical basis of discriminant analysis

Given a sample of genotypes from each breed, the object is to find functions of the genotypes that will distinguish the breeds and enable future unidentified individuals to be classified to their correct breed. Theoretically discriminant analysis can be treated as a decision problem (Welch 1939; Krzanowski and Marriott 1995)[1]. In the case of just two populations, π_1 and π_2, the individual to be classified, x, is defined as a random vector consisting of s traits or loci, having probability density functions $f_1(x)$ and $f_2(x)$ in populations one and two. An allocation rule is then defined by partitioning the p-dimensional sample space (each trait or locus occupies a separate dimension) into disjoint regions and assigning x to a particular population according to the region in which it falls. If the individual must be allocated to one or other of the two populations, and there is no "unknown" category, then the sample space is divided into just two regions, R_1 and R_2.

The probability that an individual comes from π_j and is allocated to π_i is then given by:

$$p(i/j) = \int R_i f_j(x) dx \qquad [6.1]$$

Probabilities of a correct decision can be denoted by $p(1/1)$ and $p(2/2)$, while $p(1/2)$ and $p(2/1)$ denote the probabilities of the two types of error. The concept is difficult to visualize when the distributions involved are discrete or categorical, as is the case with genotype data. It is more easily illustrated using the simple example of a continuous normally distributed trait, for example back fat, measured in two breeds.

[1] The explanatory formulae and theory given in pages 109-112 are shown in more detail in Krzanowski and Marriott (1995)

This trait may be distributed differently in the two breeds, a hypothetical example is that the first breed has a mean back fat of 6 mm with variance 4 mm, and the second breed has a mean of 25 mm with variance 6 mm. Figure 6.1 shows the two distributions, with the sample space divided into two regions (R1 and R2) by a partition at 16 mm of back fat. An unknown animal would be allocated to region R2 (breed two) if it had a back fat greater than 16 mm and to region R1 (breed one) if its back fat was less than 16 mm. The shaded areas represent the regions of error, with the light grey region being the probability of allocating an individual to breed two when it actually came from breed one ($p(2/1)$), and the dark grey region being the probability of allocating the individual to breed one when it actually came from breed two ($p(1/2)$).

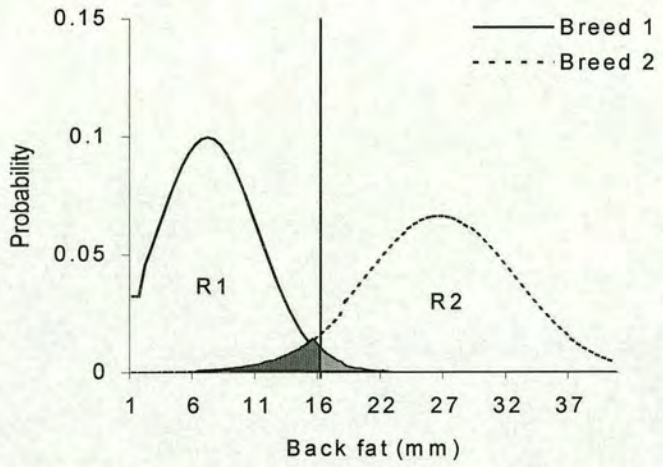

Figure 6.1. Distribution of back fat measurement in two breeds

The two types of error may not have equal consequences (i.e. one error may be more serious than the other). Unequal consequences can be accounted for by specifying two costs due to misclassification: $c(1/2)$ is the cost due to allocating a π_2 individual to π_1, and $c(2/1)$ the reverse cost. The incidence of individuals may also differ between the two populations, so prior probabilities q_1, q_2 (with $q_1+q_2 = 1$) of drawing an individual from π_1 or π_2 respectively can be specified. The probability of drawing and misclassifying an individual from π_i is then:

$$q_i p(j/i)(j \neq i)$$

[6.2]

The expected cost due to misclassifying future individuals is given by:

$$C = q_1 p(2/1)c(2/1) + q_2 p(1/2)c(1/2)$$

[6.3]

which in terms of the regions R_1 and R_2 is:

$$C = q_1 c(2/1) \int_{R2} f_1(x)dx + q_2 c(1/2) \int_{R1} f_2(x)dx$$

[6.4]

The optimal classification rule is given by the partition R_1, R_2 that minimizes the expected cost. It can be shown (Krzanowski 1988) that the optimal partition puts into R_1 all individuals x for which:

$$f_1(x)/f_2(x) \geq q_2 c(1/2)/q_1 c(2/1)$$

[6.5]

and all remaining individuals into R_2.

If the costs due to misclassification cannot be quantified then $c(1/2)$ can be set equal to $c(2/1)$. The right-hand side of equation (6.5) then becomes q_2/q_1 and the allocation rule assigns x to the population that has the greater posterior probability. If the prior probabilities are unknown then q_1 can be set equal to q_2, the right-hand side of (6.5) becomes one and x is assigned to the population in which it has the greater probability $f_i(x)$.

The concept of discriminant analysis can be extended to more than two populations. Let the prior probabilities that x comes from each of g populations be $q_1 \ldots q_g$, and the density functions of x in each of these populations be $f_1(x) \ldots f_g(x)$. A partition of the sample space into regions $R_1 \ldots R_g$ is now sought, so that x is allocated to π_i if it falls

in region R_i for i=1,...g. The probability of misclassifying an individual from π_i into π_j is now:

$$p(j/i) = \int_{Rj} f_i(x)dx$$

[6.6]

Denoting the cost of this misclassification by $c(j/i)$, the total expected cost incurred with these regions is:

$$C = \sum_{i=1}^{g} q_i \left\{ \sum_{j \neq i} c(j/i)p(j/i) \right\}$$

[6.7]

The allocation rule is given by the partition minimizing C, and it can be shown (Anderson 1984, p224) that this rule assigns x to π_k if

$$\sum_{i \neq k} q_i c(k/i) f_i(x) < \sum_{i \neq j} q_i c(j/i) f_i(x), j = 1,...., g; j \neq k$$

[6.8]

If all costs $c(j/i)$ are assumed to be equal then equation (6.8) can be simplified (Krzanowksi and Marriott 1995, p10), so that x is assigned to π_k if

$$q_j f_j(x) < q_k f_k(x), j \neq k$$

[6.9]

i.e. to the population for which its posterior probability is highest.

6.2.2 A breed allocation rule based on the multinomial distribution

The classification of individuals using genetic markers is based on the distribution of genotypes among the different breeds, the genotypes being discrete categories into which an individual can fall. Within each breed the probabilities of the genotypes can be estimated from the multinomial distribution, the parameters of each distribution being the allele frequencies in each breed. Estimates of the gene frequencies can be obtained using, for example, maximum likelihood methods on

samples from each of the breeds with these estimates then being used to formulate the allocation rule.

For locus *s*, breed *i*, category *m* (phenotype in the case of dominant loci, or genotype if the loci are co-dominant) the multinomial probabilities are p_{ism} and satisfy

$$\sum_{s=1}^{n} p_{ism} = 1,$$

where n = total number of genotypes at locus p

For example, for a simple diallelic locus, with alleles *A* and *a*, in a particular breed (where p and q are the allele frequencies for that breed) the probabilities of the three genotypes *AA*, *Aa* and *aa* (assuming Hardy-Weinberg Equilibrium) are given by p^2, $2pq$ and q^2. If the locus is dominant (as is the case with most of the cattle red cell antigen systems) then the genotypes *AA* and *Aa* cannot be distinguished, only the phenotypes; so that the probability of phenotype *A* is $p^2 + 2pq$, and that of *a* (the null phenotype in the case of blood types) is q^2. The probability of each genotype at each locus can be calculated in a similar manner, for all breeds represented in the data.

6.2.3 Assigning an unknown animal to a breed

A new animal will fall into categories $m_1,\ldots,m_s$ under the *s* loci, so the probability of this phenotype or genotype (over all loci) for breed *i* is

$$p_{im} = p_{i1m1}\, p_{i2m2} \cdots p_{isms}$$

If costs due to misclassification are assumed equal, then the rule allocates x to that breed for which $q_i p_{im}$ is greatest ($i=1 \ldots g$ breeds); p_{im} being the probability of genotype m occurring in breed i and q_i the prior probability of drawing an individual

of breed i. If equal prior probabilities are assigned to the breeds, the posterior probabilities of the breeds (with p_{im} adjusted so that it sums to one) are:

$$p(i/m) = \frac{p_{im}}{\sum_i p_{im}}$$

and x is allocated to the breed for which this probability is greatest.

This allocation rule can also be interpreted in terms of distances (Matusita 1956; Rao 1973). The allocation rule based on the multinomial distribution is the same as classifying an individual, x, to the population from which it has least distance (Krzanowski and Marriott 1995).

6.2.4. Prior probabilities

When sampling has been carried out separately from each population (the situation in most studies where a number of breeds are selected for sampling), rather than from a mixture of the breeds (which might be the case if data collected over a given time period by the Blood Typing Service were selected), there are no simple estimates of the prior probabilities (q_i) available from the data (Krzanowski and Marriott 1995). In these circumstances, unless there is extra information available, equal prior probabilities must be assumed. The analysis presented here assumes equal prior probabilities for the breeds. Additional information that might be used to provide prior probabilities is the population distribution or census sizes of breeds in various countries. In a commercial test, prior knowledge of national breed distribution or, for example, the distribution of breeds processed by a particular abattoir could be included.

6.2.5 Data

Blood type data for thirty-seven European cattle breeds (as previously described in chapters 2 and 4) were obtained from the Roslin Cattle Blood Typing Service. The markers consisted of nine independent chromosome regions, coding for the red cell antigen systems A, B, C, F, L, S, Z and the serum proteins transferrin and albumin. Within these nine regions there was a total of thirty-one loci. The majority of loci were diallelic, with the exception of the transferrin locus for which there are four alleles. The allele frequencies at the red cell antigen loci were assumed to represent the frequencies at a typical set of diallelic loci in European cattle breeds.

Allele frequencies at twenty microsatellite markers in twenty cattle populations from Europe, Africa and India were provided by David MacHugh of Trinity College Dublin (obtained from the public access ftp site: acer.gen.tcd.ie/pub/cow_microsat/). Allele frequencies were available for seven European breeds (Aberdeen Angus, Hereford, Jersey, Kerry, Charolais, Friesian and Simmental), five N'Dama populations from Africa, five other African breeds (Gobra, Maure, White Fulani, Butana, Kenana) and three Indian breeds (Hariana, Sahiwal, Tharparker). Chromosomal locations of the microsatellite markers, number of alleles and size ranges for the alleles are shown in table 6.1. These microsatellites included regions of known genes, and several anonymous sequences. The number of alleles at each locus ranged between two and twenty-two, with a mean value of 8.4. Details of how the data were collected and genotyping protocols are given in MacHugh et al (1997).

6.2.6 Methods

Comparison of microsatellite and blood type data

The analysis was based on the allele frequencies observed at the thirty-one blood type loci in thirty-seven European breeds (data set one), and the twenty microsatellite

Table 6.1 Microsatellite markers used by MacHugh et al (1997), the table shows their chromosomal location, and number and size range of the alleles.

	Locus	Gene	Chromosome	No. alleles	Size range (bp)
1	HBB	β-globin	15q22-q27	10	98-120
2	BM2113	Anonymous	2	12	123-147
3	BoLA DRP1	MHC class II β DR pseudogene	23	13	118-142
4	BoLA DR2B	MHC class II β 2 gene	23	2	144-152
5	RBP3	Retinol binding protein 3	28	6	141-153
6	PRL	Prolactin	23	3	158-164
7	ETH131	Anonymous	21	22	138-168
8	HRH1	Histamine H1 receptor	22	6	180-190
9	ILSTS014	Anonymous	19	4	128-134
10	ILSTS005	Anonymous	10	6	181-193
11	BTMICROS	Anonymous	Unassigned	13	141-187
12	HEL1	Anonymous	15	8	101-117
13	HEL5	Anonymous	21	11	161-181
14	OCAM	Opioid binding molecule	25	6	178-190
15	ETH152	Anonymous	5	11	191-211
16	ETH225	Anonymous	9	10	140-160
17	ILSTS001	Anonymous	7	11	77-97
18	RASA	RAS p21 protein activator	7q24-qter	6	182-196
19	TGLA48	Anonymous	7	4	73-79
20	TGLA116	Weaver	4	4	79-85

loci in twenty European, African and Indian breeds (data set two), as described above. The relationship between the two data sets, in terms of the information they contained about breed relationships, was tested by comparing the genetic distances (Reynolds et al 1983) between the seven European breeds (Aberdeen Angus, Charolais, Friesian, Hereford, Jersey, Kerry and Simmental) that were common to both data sets. Genetic distances between the breeds were firstly calculated from the microsatellite allele frequencies and, secondly, from the blood type allele

frequencies. The significance of the correlation between the two distance matrices was tested using Mantel's test (Manly 1986).

Data simulation

The observed allele frequencies for each of the two original data sets were used to simulate new genotypes. Initially all breeds in the original data were simulated, i.e. genotypes at 20 microsatellite loci in 20 breeds (European, African and Indian) and genotypes at 31 diallelic (blood type) loci in 37 European breeds. A direct comparison between the different marker types was then carried out using simulated data from only the seven European breeds that were common to the two original data sets (Aberdeen Angus, Charolais, Friesian, Hereford, Jersey, Kerry, Simmental). Data, based on the gene frequencies for these seven breeds, were simulated for three different situations. Microsatellite and diallelic markers were generated for the seven pure breeds, microsatellite markers only were generated for five of these breeds (Aberdeen Angus, Charolais, Friesian, Hereford and Simmental) and four crosses between the Friesian and Aberdeen Angus, Charolais, Hereford or Simmental. These particular crosses were chosen as they represent typical crosses that might be made by dairy farmers when using terminal beef sires on their non-elite cows. Genotypes were simulated for the pure breed animals assuming Hardy-Weinberg equilibrium (random mating within breeds). Crossbreed genotypes were generated by sampling one allele from one parental breed and the second from the other parental breed. The number of animals (genotypes) simulated for each breed was the same number as in the original samples, and 50 animals were generated for each crossbreed population (see MacHugh et al (1997) for number of animals sampled for microsatellite data, and chapter 4 for number of animals sampled for blood type data). The same number of animals as in the original pure breed samples were simulated in order to reflect the level of uncertainty (variance) in the original sampling process. The simulation was repeated 200 times for each data set.

Random sampling of markers

The effect of increasing the number of markers was investigated by randomly sampling markers with replacement. For example, from the original 20 microsatellite loci different numbers of markers were sampled, ranging from 1 up to 60. For the diallelic markers the number sampled ranged between 1 and 250. The sampling of markers assumes that the original markers are a random set of markers of their type, and are representative of the distribution of allele frequencies at any such loci.

Allocation of an individual to a breed

The probability of each simulated individual's genotype occurring in each breed was calculated over all markers. Assuming Hardy-Weinberg equilibrium within the pure breeds, genotype probabilities at a single locus are p_{1i}^2 for the homozygotes and $2p_{1i}p_{1j}$ for the heterozygotes. The genotype probabilities in the crossbreeds are $p_{1i}p_{2i}$ for the homozygotes and $p_{1i}p_{2j} + p_{1j}p_{2i}$ for the heterozygotes. Where p_{1i} and p_{2i} are the frequencies of allele i in breeds 1 and 2, and p_{1j} and p_{2j} are the frequencies of allele j in breeds 1 and 2, where $i \neq j$. The genotype probability over all m markers is given by $\prod_{i=1}^{m} p_i$, and the individual is allocated to the breed for which it has the highest probability (see sections 6.2.1 to 6.2.3 for further details).

Estimation of the error rate

The efficiency of the different marker numbers for discriminating among the breeds was measured by calculating the proportion of animals that were misclassified for each set of markers. The overall proportion of animals misclassified (error rate) was calculated as the total number of animals allocated to the wrong breeds divided by the total number of animals to be allocated.

Two types of error can be determined. Type I error, or the proportion of individuals of one particular breed that are allocated to another breed (really breed A but allocated to breed B), and type II error, or the proportion of individuals that are allocated to one breed but are really of another breed (allocated to breed A but really a member of breed B). For any individual breed the two types of error may not be equal, however, across all breeds being compared the two error rates are symmetric (equal to one another). Both types of error were calculated for each individual breed: type I error as the number of animals of the breed that were allocated to other breeds divided by the number of animals that were really members of that breed, type II error as the number of animals allocated to the breed that were actually of another breed divided by the total number of animals that were allocated to the breed.

Selection of microsatellite markers

The efficiency of individual microsatellite markers for discriminating among breeds was investigated by estimating the error rate for each marker separately. Markers were then selected on their individual error rates, the best being those with the lowest individual error rates. The best markers were combined, markers with lowest error rates being added first, and the resulting combined error rates were compared with those obtained when markers were randomly sampled. Marker selection was carried out with the microsatellite markers only, as they were more variable in numbers of alleles and average heterozygosity than the diallelic markers, allowing any relationship between these measures and error rate to be seen more clearly.

In addition, the relationship between marker variability and error rates was investigated by calculating the average heterozygosity, number of alleles observed (see chapter 4 section 4.2.4 for method) and F_{ST} (see chapter 1 section 1.5.1 for formula) for each individual marker. The average heterozygosity, number of alleles observed and average genetic distance from other breeds was also calculated for each of the breeds in the data sets, in order to establish any relationship between these statistics and the ease with which any particular breed could be identified.

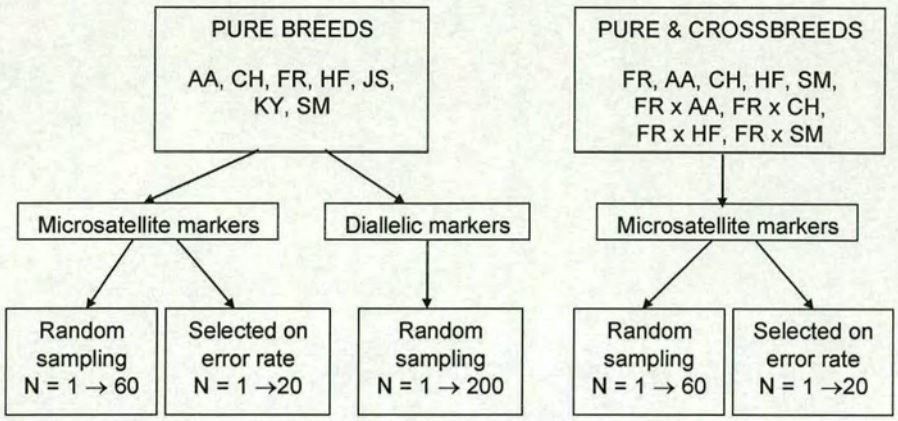

Figure 6.2a Diagram showing the structure of the simulated data sets in terms of the breeds included, marker type and whether markers were randomly sampled or selected (N is the range of the number of markers sampled).

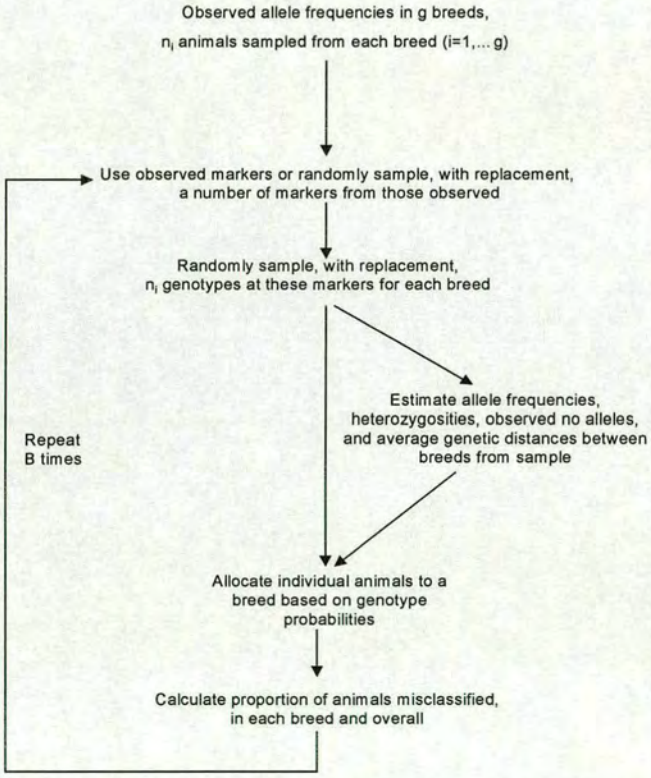

Figure 6.2b Flow diagram of the strategy for estimation of error rates, and measures of genetic variation (heterozygosity, observed number alleles and average genetic distance between breeds).

Figure 6.2a shows, in diagrammatic form, the various data sets that were simulated and whether markers were randomly sampled or selected. For each data set 200 replicates were generated, i.e. a new data set was simulated, the allocation procedure carried out, and the error rate and other statistics estimated. This was carried out 200 times, and the mean error rate, average heterozygosity, number of alleles observed and average genetic distance or F_{ST} calculated over all the replicates. The procedure is illustrated in figure 6.2b.

6.3 Results

6.3.1 Comparison of diallelic markers with microsatellite markers

The microsatellite and blood type data gave related estimates of genetic distances among the seven European breeds they had in common. The correlation between the two distance matrices was 0.56, which was significantly different from zero ($p < 0.05$), and shows that the different marker types give estimates of the genetic distances among breeds that are relatively similar. Figure 6.3 shows the relationship between the genetic distances estimated using either microsatellite or blood type data for the seven European cattle breeds. The positive correlation between the two estimates can be clearly seen. Outlying values are the distance between the Charolais and Friesian (where the distance estimated from the microsatellite data is smaller than that estimated from the blood type data), and the distances between the Aberdeen Angus and the Kerry, Simmental or Friesian (where the distances estimated from the microsatellite data are larger than those estimated from the blood type data).

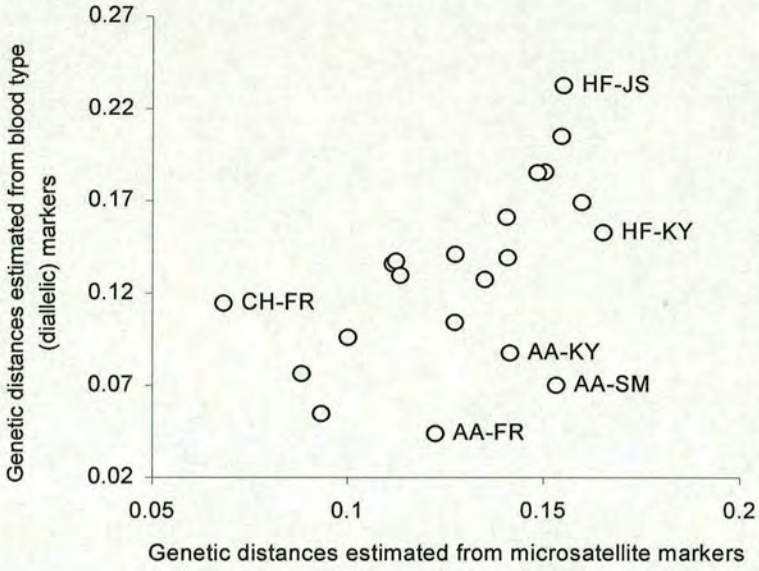

Figure 6.3 Relationship between genetic distances among 7 European breeds estimated using microsatellite markers or diallelic blood type markers. Outlying points are labelled with the two breeds to which the distance value relates.

The mean proportions, over the 200 replicates, of animals correctly identified in each breed using the microsatellite markers are shown in table 6.2, and using diallelic markers in table 6.3. Figure 6.4a shows the relationship between the mean error rate and the number of markers for the microsatellite data. The graph shows that the error rates are higher when all 20 breeds are included for identification compared with only the seven European breeds. Error rates below 5% can be achieved with 11 or more markers if the comparison is between the seven European breeds, and with 18 or more markers for all 20 breeds.

Figure 6.4b shows the relationship between the mean error rate and the number of markers for the diallelic markers. When only the seven European breeds are compared, then around 65 markers are required to achieve error rates below 5%. If all 37 breeds are compared then about 100 markers are required to achieve the same accuracy. It is clear that the number of breeds to be compared has an effect on the number of markers required to achieve a given error rate, the more breeds to be compared the more markers are required.

Table 6.2a Error rate over all breeds and proportion of animals correctly classified in each breed (mean over 200 bootstrap replicates) with increasing numbers of randomly sampled microsatellite markers (comparing 20 European, African and Indian breeds).

No. markers	Overall error rate	Proportion of animals correctly classified in each breed								
		AA	HF	JS	KY	CH	FR	SM	ND1	ND2
5	0.348	0.779	0.827	0.816	0.797	0.711	0.712	0.771	0.398	0.613
10	0.134	0.959	0.961	0.967	0.965	0.925	0.922	0.944	0.696	0.816
15	0.069	0.987	0.992	0.991	0.986	0.974	0.977	0.982	0.824	0.888
20	0.037	0.997	0.998	0.998	0.997	0.995	0.995	0.997	0.910	0.939
30	0.013	1.000	1.000	1.000	0.999	0.998	1.000	1.000	0.966	0.977
40	0.003	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	0.992
50	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.996
60	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998

Table 6.2a. Error rate over all breeds and proportion of animals correctly classified in each breed (mean over 200 bootstrap replicates) with increasing numbers of randomly sampled microsatellite markers (comparing 20 European, African and Indian breeds) – continued.

No. markers	Proportion of animals correctly classified in each breed										
	ND3	ND4	ND5	GO	ME	WF	BU	KE	HA	SW	TH
5	0.521	0.616	0.467	0.559	0.515	0.610	0.736	0.682	0.865	0.849	0.910
10	0.759	0.845	0.720	0.813	0.814	0.859	0.937	0.917	0.982	0.976	0.993
15	0.846	0.927	0.829	0.913	0.895	0.938	0.979	0.975	0.998	0.996	0.998
20	0.900	0.969	0.901	0.955	0.937	0.968	0.994	0.989	1.000	0.998	1.000
30	0.962	0.990	0.961	0.985	0.980	0.993	1.000	0.999	1.000	0.999	1.000
40	0.987	1.000	0.991	0.995	0.996	0.998	1.000	1.000	1.000	1.000	1.000
50	0.995	1.000	0.996	0.999	0.998	1.000	1.000	1.000	1.000	1.000	1.000
60	0.998	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000

Table 6.2b. Error rate over all breeds and proportion of animals correctly classified in each breed (mean over 200 bootstrap replicates) with increasing numbers of randomly sampled microsatellite markers (comparing 7 European breeds).

No. markers	Overall error rate	Proportion of animals correctly classified in each breed						
		AA	CH	FR	HF	JS	KY	SM
5	0.223	0.802	0.822	0.812	0.790	0.746	0.714	0.760
10	0.054	0.953	0.923	0.926	0.958	0.953	0.957	0.945
15	0.013	0.990	0.977	0.980	0.991	0.991	0.991	0.985
20	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
30	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6.3a. Error rates over all breeds and proportion of animals correctly classified in each breed (mean over 100 bootstrap replicates) with increasing numbers of randomly sampled diallelic markers (37 European breeds)

		Proportion of animals correctly classified in each breed							
No. markers	Overall error	AA	AY	BB	BG	BA	BW	BS	CH
25	0.460	0.234	0.309	0.314	0.348	0.514	0.636	0.655	0.555
50	0.203	0.537	0.643	0.617	0.633	0.811	0.881	0.908	0.848
75	0.098	0.731	0.803	0.786	0.809	0.927	0.957	0.972	0.939
100	0.050	0.820	0.840	0.850	0.910	0.970	0.980	0.990	0.960
125	0.027	0.924	0.946	0.938	0.947	0.986	0.996	0.998	0.992
150	0.015	0.961	0.965	0.968	0.973	0.994	0.998	1.000	0.997
175	0.010	0.977	0.983	0.983	0.988	0.998	0.999	1.000	0.999
200	0.009	0.987	0.987	0.990	0.991	0.999	1.000	1.000	1.000

	Proportion of animals correctly classified in each breed								
No. markers	DX	HO	GA	GB	GL	GU	HF	PH	HL
25	0.366	0.299	0.359	0.687	0.773	0.475	0.616	0.501	0.625
50	0.684	0.616	0.669	0.909	0.947	0.789	0.759	0.713	0.873
75	0.837	0.781	0.824	0.973	0.990	0.919	0.826	0.807	0.949
100	0.920	0.910	0.900	0.996	0.990	0.980	0.830	0.810	0.980
125	0.965	0.928	0.954	0.997	1.000	0.989	0.889	0.885	0.994
150	0.983	0.959	0.974	0.999	1.000	0.995	0.908	0.904	0.998
175	0.991	0.975	0.986	0.999	1.000	0.998	0.929	0.922	0.999
200	0.996	0.986	0.993	1.000	1.000	1.000	0.936	0.935	1.000

	Proportion of animals correctly classified in each breed								
No. markers	JS	KY	LM	LO	LR	MA	MC	MR	MU
25	0.821	0.641	0.466	0.810	0.823	0.424	0.754	0.285	0.460
50	0.969	0.871	0.773	0.959	0.969	0.747	0.946	0.580	0.774
75	0.993	0.951	0.901	0.992	0.990	0.892	0.986	0.761	0.895
100	1.000	0.990	0.990	0.990	1.000	0.980	1.000	0.890	0.950
125	1.000	0.992	0.984	1.000	1.000	0.983	1.000	0.928	0.979
150	1.000	0.998	0.991	1.000	1.000	0.993	1.000	0.962	0.991
175	1.000	1.000	0.997	1.000	1.000	0.998	1.000	0.975	0.995
200	1.000	1.000	0.998	1.000	1.000	0.999	1.000	0.984	0.997

	Proportion of animals correctly classified in each breed							
No. markers	RM	SA	SH	SM	SD	SX	WB	WP
25	0.762	0.625	0.427	0.331	0.528	0.677	0.309	0.621
50	0.945	0.861	0.726	0.657	0.843	0.904	0.647	0.861
75	0.990	0.942	0.883	0.830	0.939	0.976	0.828	0.947
100	1.000	0.970	0.880	0.890	0.990	0.990	0.890	0.991
125	1.000	0.990	0.975	0.961	0.989	0.997	0.953	0.992
150	1.000	0.994	0.989	0.981	0.996	0.999	0.979	0.996
175	1.000	0.998	0.994	0.990	0.998	1.000	0.987	0.999
200	1.000	0.999	0.997	0.996	1.000	1.000	0.995	0.999

Table 6.3b Error rate over all breeds and proportion of animals correctly classified in each breed (mean over 200 bootstrap replicates) with increasing numbers of randomly sampled diallelic markers (7 European breeds compared).

		Proportion of animals correctly classified in each breed						
No markers	Overall error	AA	CH	HO	HF	JS	KY	SM
25	0.264	0.664	0.814	0.690	0.924	0.936	0.867	0.696
50	0.103	0.861	0.945	0.871	0.992	0.996	0.964	0.885
75	0.024	0.947	0.985	0.953	0.999	1.000	0.988	0.961
100	0.011	0.975	0.994	0.977	1.000	1.000	0.996	0.983
125	0.003	0.990	0.999	0.991	1.000	1.000	0.999	0.994
150	0.001	0.994	0.999	0.995	1.000	1.000	1.000	0.996
175	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

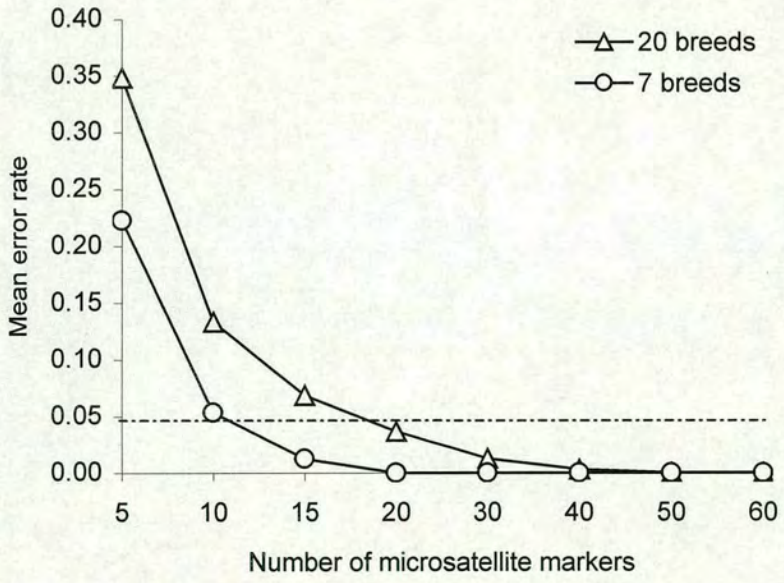

Figure 6.4a. Relationship between number of randomly sampled microsatellite markers and the mean error rate, when either 20 (European, African and Indian) or 7 (European) breeds are compared.

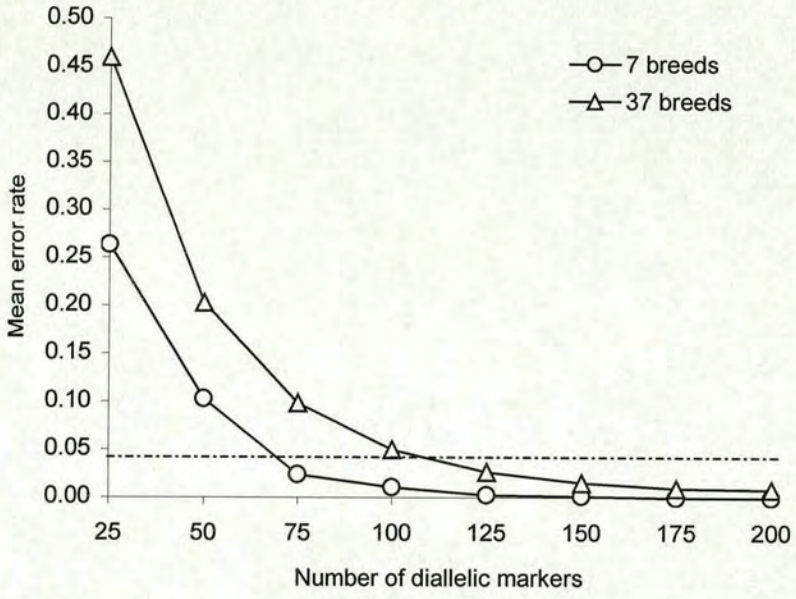

Figure 6.4b. Relationship between number of randomly sampled diallelic markers and the mean error rate, when either 37 or 7 European breeds are compared.

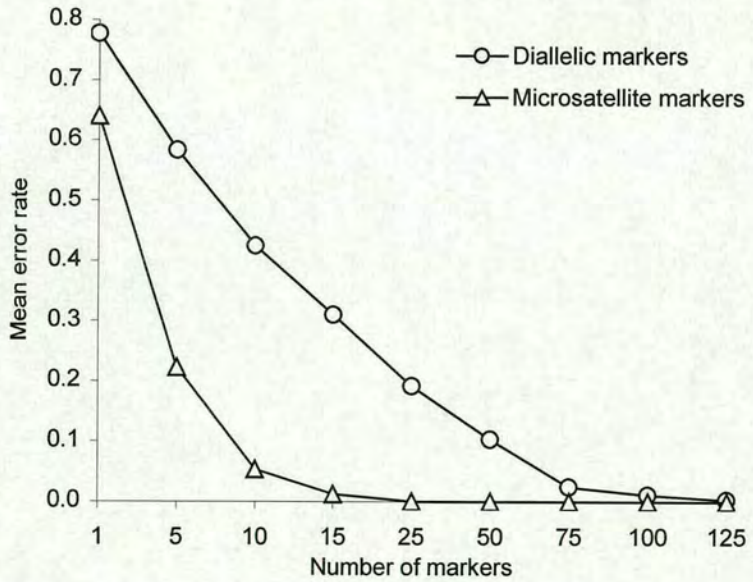

Figure 6.5 Efficiency of microsatellite markers compared with diallelic markers for discriminating among 7 European breeds (markers were randomly sampled)

Figure 6.5 shows the microsatellite and diallelic markers compared directly, when used to discriminate among seven European breeds. Approximately ten microsatellite markers are required to achieve a 5% error rate compared with about sixty-five diallelic markers. This suggests that about six diallelic markers are

required for every one microsatellite marker to reach the same power of discrimination.

6.3.2 Selection of individual microsatellite markers

Table 6.4a shows the average heterozygosity, number of alleles observed, proportion of animals correctly classified and F_{ST} for each individual microsatellite marker (mean values across all breeds). Table 6.4b shows similar statistics for each individual breed (mean values over all markers), with the average genetic distance of a breed from other breeds shown rather than F_{ST}. The relationships between average heterozygosity, average number of alleles observed, average genetic distance between breeds and error rate for each of the markers are shown in figure 6.6 (i.e. this figure relates to table 6.4a) The relationships between the same parameters, but this time for each breed, are shown in figure 6.7 (this figure relates to table 6.4b). Figure 6.6 indicates that the markers that are most efficient at distinguishing between breeds (have the lowest error rate) are the markers with highest heterozygosities, and the greatest number of alleles observed (across all breeds). However, there is no clear relationship between F_{ST} and error rate. Conversely, figure 6.7 shows that breeds that are most easily identified are those with lowest heterozygosities, smallest number of alleles, and that are genetically most distant from other breeds.

The best markers for discriminating between the breeds can be selected on the basis of their individual error rates. Figure 6.8 shows the effect of adding individual microsatellite markers one by one, with markers ranked according to their individual error rate i.e. markers with lowest error rates are added first. It can be seen that, using the microsatellite markers to discriminate between twenty breeds, the lowest combined error rate is achieved with approximately 11-12 markers. The addition of the remaining eight or nine markers does not significantly affect the error rate. If the markers are used to discriminate between only the seven European breeds, then error rates of less than 5% can be achieved with just 5-6 markers.

Table 6.4a Average heterozygosity, number of alleles observed, F_{ST} and error rate over all twenty breeds for each microsatellite marker. Ranking is on error rate, in ascending order (i.e. marker with lowest error rate is ranked one).

Marker	Average heterozygosity	No alleles	Fst	Error	Rank (error)
1	0.673	6	0.131	0.750	5
2	0.734	6	0.164	0.664	1
3	0.790	8	0.064	0.792	9
4	0.268	2	0.078	0.949	20
5	0.464	3	0.205	0.861	15
6	0.198	2	0.132	0.895	19
7	0.763	8	0.133	0.670	2
8	0.403	3	0.218	0.833	11
9	0.328	3	0.166	0.870	17
10	0.512	3	0.224	0.844	12
11	0.626	5	0.234	0.703	3
12	0.637	5	0.240	0.719	4
13	0.690	6	0.174	0.777	8
14	0.333	3	0.359	0.852	13
15	0.513	4	0.308	0.759	6
16	0.652	5	0.201	0.765	7
17	0.444	4	0.295	0.801	10
18	0.517	3	0.204	0.872	18
19	0.507	3	0.214	0.865	16
20	0.527	3	0.222	0.854	14

Table 6.4b Average heterozygosity, number of alleles observed, average genetic distance from other breeds, and error over all twenty microsatellite markers for each breed. Ranking is on error rate, in ascending order (i.e. breed with lowest error rate is ranked one).

Breed	Average heterozygosity	No alleles	Average genetic distance from other breeds	Error	Rank (error)
AA	0.457	3	0.211	0.683	5
HF	0.445	4	0.200	0.709	7
JS	0.421	3	0.208	0.637	3
KY	0.462	3	0.203	0.709	6
CH	0.535	4	0.162	0.792	10
FR	0.531	5	0.169	0.823	14
SM	0.468	4	0.192	0.809	12
ND1	0.573	5	0.135	0.968	20
ND2	0.494	4	0.174	0.752	8
ND3	0.524	4	0.163	0.848	15
ND4	0.527	4	0.162	0.820	13
ND5	0.540	5	0.150	0.907	17
GO	0.617	5	0.140	0.932	18
ME	0.645	6	0.142	0.933	19
WF	0.619	5	0.152	0.861	16
BU	0.581	4	0.185	0.788	9
KE	0.597	4	0.189	0.800	11
HA	0.497	4	0.261	0.627	2
SW	0.527	4	0.243	0.671	4
TH	0.520	3	0.258	0.502	1

The twelve individual microsatellite markers with the lowest error rates when discriminating among 20 European, African and Indian breeds were 2, 7, 11, 12, 1, 15, 16, 13, 3, 17, 8, and 10 (see table 6.1 for marker details). These are all anonymous microsatellite loci, that are not known to form part of functional genes, with the exception of locus 1 (β globin gene), locus 3 (MHC class II β DR pseudogene) and locus 8 (histamine H1 receptor gene). Average number of alleles observed for these twelve markers was 5.08, mean heterozygosity was 0.55, and average genetic distance between breeds was 0.2.

These results indicate that the most powerful markers for discriminating between the breeds studied are microsatellite markers, particularly those that have high average heterozygosities, and observed numbers of alleles (mean across all the breeds). If markers are specifically selected as being the most discriminatory (having the lowest individual error rates) then the number of markers required for accurate discrimination between breeds can be reduced. It is, however, likely that the best markers for discriminating between one set of breeds will not necessarily be the best for another set. The problem is illustrated in figure 6.8 where the first marker, which is the most discriminatory when 20 breeds are being compared, has a higher error rate when only 7 European breeds are being compared. In practice, some kind of optimization procedure could be used to select the best markers for the problem at hand.

Figure 6.9 shows the difference between using randomly sampled markers compared with selecting markers on their individual error rates. To discriminate among the seven breeds with an error rate of 5% requires about 10 randomly sampled markers compared with 5-6 selected markers.

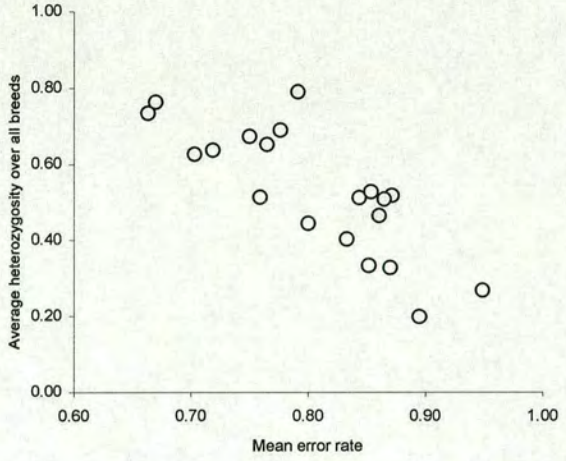

Figure 6.6a. Average heterozygosity plotted against the mean error rate for each of 20 microsatellite markers

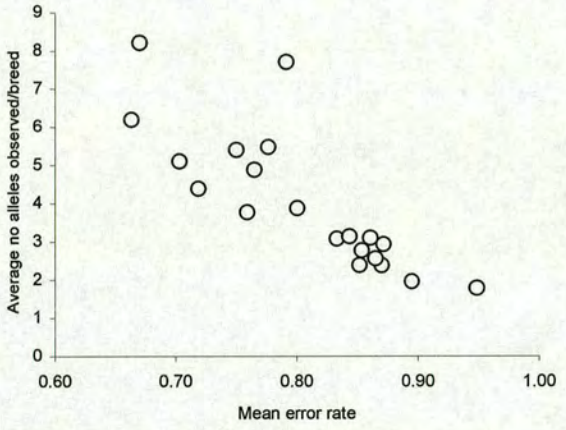

Figure 6.6b. Average number of alleles observed per breed plotted against the mean error rate for each of 20 microsatellite markers

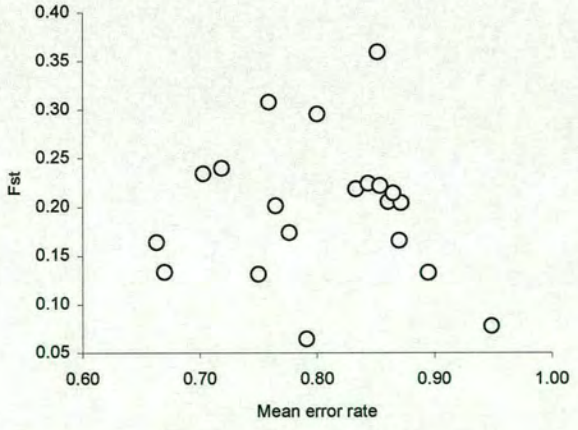

Figure 6.6c. F_{ST} plotted against mean error rate for each of 20 microsatellite markers.

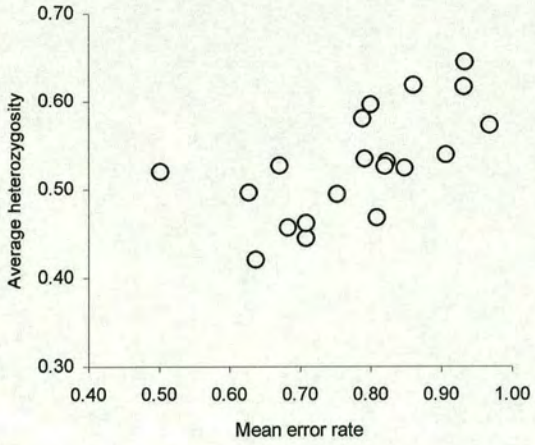

Figure 6.7a. Average heterozygosity within breeds plotted against the mean error rate for each of 20 breeds

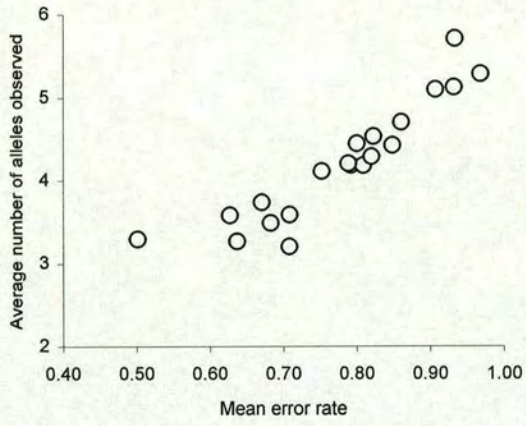

Figure 6.7b. Average number of alleles observed within breeds plotted against the mean error rate for each of 20 breeds

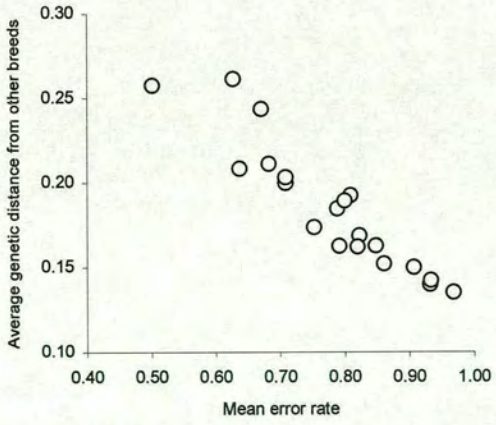

Figure 6.7c. Average genetic distance of breed from other breeds plotted against the mean error rate for each of 20 breeds

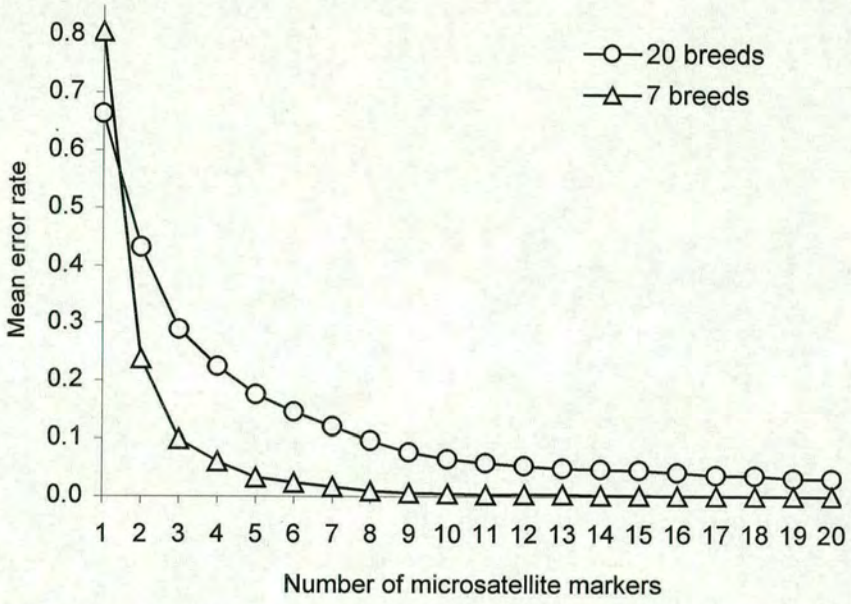

Figure 6.8. Effect on error rate of using selected microsatellite markers (with markers added in order of lowest individual error rates) when comparing 20 European, African and Indian breeds or 7 European breeds.

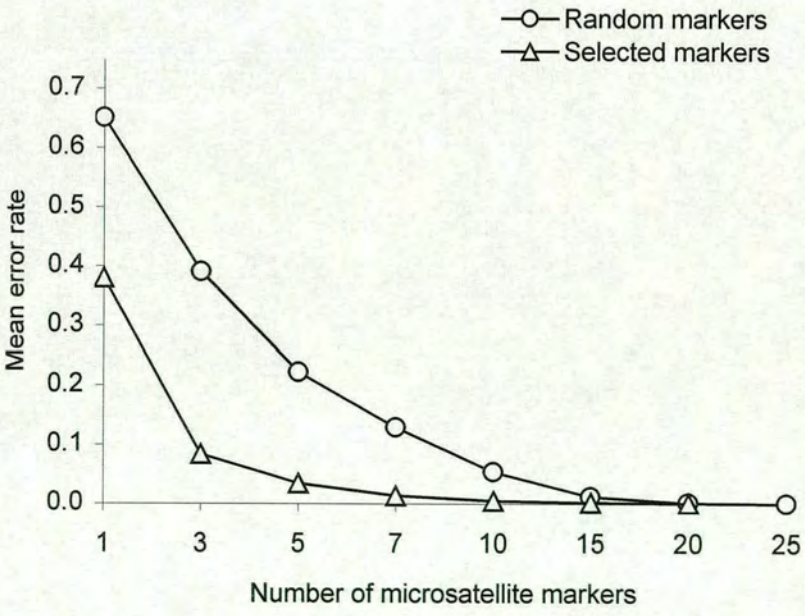

Figure 6.9 Comparison between randomly sampled microsatellite markers and markers selected on individual error rates, when used to discriminate among 7 pure European breeds

6.3.3 Discrimination among pure and crossbreeds

First crosses between breeds are closely related to both parental breeds, and this makes the problem of discrimination more difficult. Table 6.5 shows both type I and type II error when the best five selected microsatellite markers (markers selected on individual error rate) were used to discriminate between five pure (Aberdeen Angus, Charolais, Friesian, Hereford, Simmental) and four crossbreeds (Friesian crossed with Aberdeen Angus, Charolais, Hereford or Simmental). Table 6.6 shows type I and type II error rates when the best ten markers were used to distinguish between the same pure and crossbreeds. The highest type I errors are seen in the first cross Friesians, i.e. a reasonably high proportion of these animals, 30-40% with five markers or 15-30% with ten markers, are mistaken for another breed (usually one of the parental breeds or another of the Friesian crosses). Type II error is highest for the pure bred Friesian where 36% (five markers) or 24% (ten markers) of animals allocated to Friesian are from other breeds, these animals are most likely to be first cross Friesians. Type II error is also high for the crossbred Friesians, 20-30% (five markers) or 13-20% (ten markers). This is because the animals from the pure bred parental breeds or other Friesian crosses are confused with them.

Figure 6.10 illustrates the use of randomly sampled markers against selected markers for discriminating between pure and crossbreed animals. About 30 randomly sampled markers were sufficient to achieve error rates of 5%. The selection of markers was restricted by the fact that in total only 20 markers were available. The line levels off at around 12% error at seven or more markers. It is possible that given markers that were all equally as discriminatory as the best five then error rates of 5% or less could be achieved with no more than 10-15 markers, at least for the crossbreeds chosen in this example.

Table 6.5a Each row shows proportion of animals of the (row) breed that were allocated to each category. The proportion of animals allocated to breeds other than their true breed is the type I error. Results shown were obtained by using the five best microsatellite markers, selected for lowest individual error rates, to discriminate among five pure European breeds and four Friesian crosses.

	Allocated breed									
True breed	AA	CH	FR	HF	SM	FR x AA	FR x CH	FR x HF	FR x SM	Type I error
AA	0.920	0.004	0.002	0.007	0.001	0.058	0.003	0.004	0.001	0.080
CH	0.006	0.854	0.005	0.009	0.012	0.010	0.084	0.006	0.014	0.146
FR	0.003	0.003	0.808	0.001	0.006	0.081	0.049	0.026	0.024	0.192
HF	0.010	0.009	0.001	0.904	0.001	0.006	0.004	0.064	0.001	0.096
SM	0.000	0.010	0.003	0.001	0.918	0.003	0.008	0.002	0.054	0.082
FR x AA	0.096	0.008	0.086	0.006	0.004	0.701	0.046	0.045	0.009	0.299
FR x CH	0.007	0.115	0.093	0.004	0.016	0.063	0.590	0.043	0.068	0.410
FR x HF	0.010	0.012	0.056	0.079	0.003	0.073	0.060	0.696	0.013	0.304
FR x SM	0.003	0.013	0.114	0.001	0.121	0.029	0.055	0.014	0.650	0.350

Table 6.5b Each column shows proportion of animals from each breed that were allocated in each (column) category. The proportion of animals allocated to a breed, which were really of another breed, is type II error. Results shown were obtained by using the five best microsatellite markers, selected for lowest individual error rates to discriminate among five pure European breeds and four Friesian crosses.

	Allocated breed								
True breed	AA	CH	FR	HF	SM	FR x AA	FR x CH	FR x HF	FR x SM
AA	0.825	0.004	0.001	0.006	0.000	0.039	0.003	0.003	0.000
CH	0.006	0.788	0.003	0.009	0.010	0.008	0.071	0.005	0.012
FR	0.003	0.003	0.644	0.001	0.006	0.066	0.045	0.024	0.024
HF	0.010	0.008	0.001	0.856	0.001	0.004	0.003	0.050	0.001
SM	0.000	0.009	0.002	0.001	0.806	0.003	0.007	0.002	0.048
FR x AA	0.130	0.010	0.085	0.009	0.005	0.713	0.053	0.052	0.011
FR x CH	0.010	0.148	0.093	0.006	0.020	0.064	0.686	0.050	0.084
FR x HF	0.013	0.015	0.055	0.109	0.004	0.075	0.069	0.799	0.016
FR x SM	0.004	0.017	0.114	0.002	0.147	0.030	0.064	0.016	0.804
Type II error	0.175	0.212	0.356	0.144	0.194	0.287	0.314	0.201	0.196

Table 6.6a Each row shows proportion of animals of the (row) breed that were allocated to each category. The proportion of animals allocated to breeds other than their true breed is the type I error. Results shown were obtained by using the ten best microsatellite markers, selected for lowest individual error rates, to discriminate among five pure European breeds and four Friesian crosses.

	Allocated breed									
True breed	AA	CH	FR	HF	SM	FR x AA	FR x CH	FR x HF	FR x SM	Type I error
AA	0.980	0.001	0.000	0.002	0.000	0.017	0.000	0.001	0.000	0.020
CH	0.001	0.929	0.001	0.001	0.006	0.004	0.054	0.001	0.004	0.071
FR	0.000	0.000	0.856	0.000	0.001	0.045	0.047	0.022	0.029	0.144
HF	0.001	0.000	0.000	0.972	0.000	0.001	0.000	0.025	0.000	0.028
SM	0.000	0.004	0.001	0.000	0.952	0.001	0.003	0.002	0.038	0.048
FR x AA	0.040	0.002	0.043	0.002	0.001	0.842	0.026	0.035	0.009	0.157
FR x CH	0.001	0.080	0.065	0.001	0.006	0.038	0.723	0.030	0.056	0.277
FR x HF	0.002	0.004	0.034	0.039	0.004	0.048	0.027	0.829	0.014	0.171
FR x SM	0.000	0.004	0.068	0.000	0.074	0.021	0.048	0.017	0.767	0.232

Table 6.6b Each column shows proportion of animals from each breed that were allocated in each (column) category. The proportion of animals allocated to a breed, which were really of another breed, is type II error. Results shown were obtained by using the ten best microsatellite markers, selected for lowest individual error rates, to discriminate among five pure European breeds and four Friesian crosses.

	Allocated breed								
True breed	AA	CH	FR	HF	SM	FR x AA	FR x CH	FR x HF	FR x SM
AA	0.935	0.001	0.000	0.002	0.000	0.011	0.000	0.001	0.000
CH	0.001	0.876	0.001	0.001	0.005	0.003	0.043	0.001	0.003
FR	0.000	0.000	0.765	0.000	0.001	0.036	0.041	0.019	0.025
HF	0.001	0.000	0.000	0.939	0.000	0.001	0.000	0.018	0.000
SM	0.000	0.003	0.001	0.000	0.885	0.001	0.002	0.001	0.030
FR x AA	0.058	0.003	0.048	0.002	0.001	0.841	0.028	0.037	0.010
FR x CH	0.001	0.105	0.072	0.001	0.008	0.038	0.802	0.031	0.062
FR x HF	0.003	0.005	0.038	0.055	0.005	0.048	0.030	0.874	0.016
FR x SM	0.000	0.006	0.076	0.000	0.095	0.021	0.053	0.018	0.853
Type II error	0.065	0.124	0.235	0.061	0.115	0.159	0.198	0.126	0.147

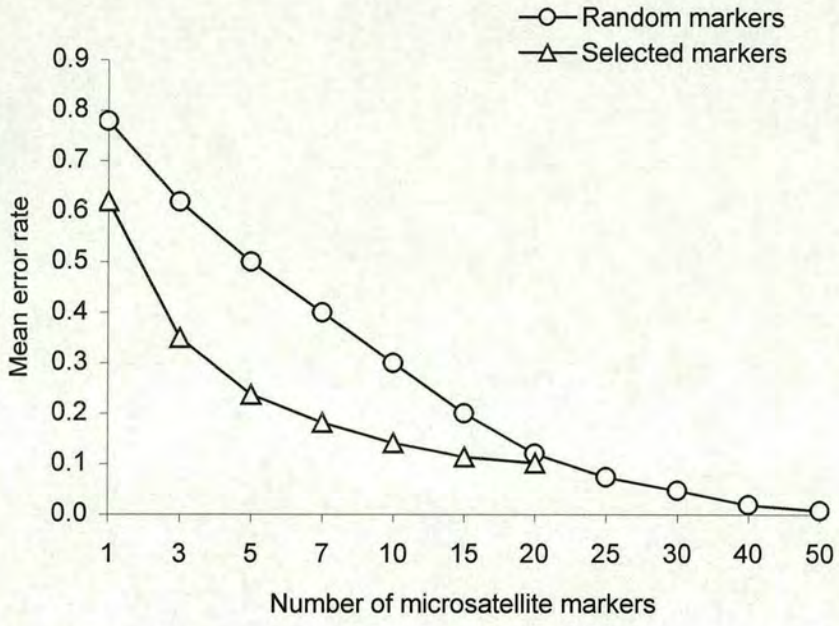

Figure 6.10 Comparison between randomly sampled markers and markers selected on individual error rates, when used to discriminate among 5 pure breeds and 4 Friesian crosses

6.4 Discussion

The results presented in this chapter show that microsatellite markers are more powerful than the diallelic markers for distinguishing between breeds, at least when the diallelic marker frequencies are similar to the allele frequencies observed at blood type loci. Error rates below 5% can be achieved with 11-18 randomly sampled microsatellites (depending on the breeds being compared) against 65-100 randomly sampled diallelic markers. If the most discriminatory markers are selected then the number of markers required to achieve the same error rates can be reduced, by about half. The most powerful markers for breed discrimination are those with high average heterozygosities, and with greater mean numbers of observed alleles (across all breeds). A characteristic of microsatellite markers is that they tend to have high heterozygosities and numbers of alleles, due to their higher rate of mutation compared with protein loci. The between population variance in gene frequencies, estimated by F_{ST}, does not appear to have any clear relationship with the error rate for individual microsatellite markers, which suggests that any concern that markers with

high diversity (high average heterozygosity across breeds, which tends to be negatively correlated with F_{ST}) (Bowcock et al 1994) will be less informative for distinguishing between populations is unfounded.

A feature of the microsatellite data set analysed was that it contained both *Bos taurus* and *Bos indicus* breeds which have been shown to be deeply divergent (Loftus et al 1994). Some of the microsatellite loci were found to have alleles that were diagnostic for *Bos indicus* (MacHugh 1996) and consequently they had higher values of F_{ST} than other markers. This, however, had no bearing on the error rate for the markers because the error rate is a measure of the marker efficiency for discriminating among all the breeds in the data set. For example, even if the marker is diagnostic for *Bos indicus* unless it also distinguishes among individual *Bos indicus* breeds or populations then the error rate for the marker is likely to be high. It would only be low if trying to discriminate between a single *Bos taurus* breed and a single *Bos indicus* breed.

The ease with which different breeds can be identified is related to not just the variation within breeds (heterozygosity) and number of alleles observed, but also to the average genetic distance of a breed from the other breeds. Breeds with low heterozygosities, smaller numbers of observed alleles and greater distances away from other breeds are more easily recognised. For example, breeds with small effective population sizes or that have been through population bottlenecks (e.g. rare breeds) are more likely to be accurately identified using genetic markers. The number of breeds that are being compared also affects the error rates obtained with a given set of markers. The higher the number of breeds being compared, the more markers are required, particularly if the breeds are closely related. However, if enough markers are used then this effect can be eliminated, although the relationship between number of breeds and number of markers required is complex since it depends on the genetic distances between the breeds being compared. For the twenty breeds represented in the original microsatellite data set, thirty or more randomly sampled markers are sufficient to reduce the error rates to almost zero. When only

seven European breeds are compared the error rate is less than 2% using fifteen markers, however if some crossbreeds are included then more than forty markers are required to achieve the same error rate. The precise number of markers required to achieve a given error rate will depend on the number and type of populations that are to be compared.

The most powerful markers for distinguishing among populations are those that are fixed for different alleles in different breeds (breed specific alleles). Traditionally, European breeds have been distinguished by phenotypic differences such as coat colour, presence or absence of horns and other traits such as double muscling. These traits tend to be fixed within breeds and to differ markedly between breeds. Recent advances in bovine genome mapping have led to the identification of the DNA regions controlling some of these traits. Markers are now available for the extension locus, which plays a role in determining red/black coat colour (Joerg et al 1996; Adalsteinsson et al 1995). Other loci that have been mapped are roan (Charlier et al 1996), the poll gene (Georges et al 1993; Brenneman et al 1996) and double muscling (Grobet et al 1997). It has been suggested that these markers may be useful in a DNA test of breed (Georges and Andersson 1996).

At least ten major, and many more minor, loci control coat colour in mammals (Searle 1968; Jackson 1994). However, there is currently limited evidence to support the assertion that different alleles at these loci are fixed in different cattle breeds. Adalsteinsson et al (1995) studied coat colour in the Icelandic breed of cattle, a breed in which a variety of colours occur, with no deliberate selection for or against any particular colour. They found three alleles at the extension locus (E^d, E^+ and e) present at frequencies of 0.083, 0.343 and 0.574, and two alleles at the agouti locus (A^+ and a) present at frequencies of 0.925 and 0.075. Interaction between loci plays an important role in determining the phenotype of an animal, and simply knowing the genotype at a single locus may be insufficient. For example, presence of the E^+ allele at the extension locus generally results in a black phenotype, with red animals being homozygous for the e allele. However, Klungland et al (1995) found red Norwegian

cattle with all three genotypes (E^+/E^+, E^+/e and e/e) at the locus. They postulated that the E^+ allele was inhibited in these animals by a dominant allele at the agouti locus (cf. dominant red allele at the agouti locus in mice).

Allele frequencies at the coat colour loci in other European cattle breeds have not been estimated. Many breeds that have a single predominant colour (e.g. black in Aberdeen Angus, or Holstein-Friesian) may also have other colour alleles segregating at low frequencies. The Holstein breed is divided into red and black subpopulations, with gene flow between the two occurring through rare black carriers of the recessive red allele (Joerg et al 1996). Occasionally, red spotted or red-roan spotted animals are encountered in the Belgian Blue, a predominantly grey and white breed, suggesting that there is still segregation of the red allele in this breed albeit at low frequency (Charlier et al 1996). In many breeds different colour variants are accepted by breeders, in other breeds the effect of modifier loci may obscure variation at other loci. For example, it is conceivable that the Charolais breed may be segregating for both black and red alleles at the extension locus, but the phenotype is cream or white due to the effect of the dilution locus on the hair colour. Many genes (e.g. polling) have been introgressed from one breed to others, and this is likely to continue, with the result that these types of marker will no longer be useful for distinguishing among breeds. As the effects of these genes are clearly visible, breeders are able to directly select for a desired phenotype. Markers that reflect "hidden" genetic variation may therefore prove more powerful for distinguishing between breeds, since they are not so easily manipulated by breeders.

The existence of markers with breed specific alleles would dramatically increase the power of breed identification. Only a small number of loci with breed specific alleles, for example four or five, would require to be genotyped in order to obtain very accurate discrimination between breeds. In order to isolate such markers large numbers of breeds may need to be screened for several hundred loci. In the mean time, the microsatellite markers that are currently available could provide an adequate means of discriminating among cattle populations, particularly if marker

combinations are selected in an optimal way i.e. for lowest error rates for the problem at hand.

Practical implementation of a test of breed identity may require several further refinements to the techniques described here. The use of codominant markers, such as microsatellites, would allow genotype frequencies to be used directly rather than predicting genotype frequencies from the allele frequencies which requires the assumption of random mating. A potential problem in doing this is that sparseness of the data may be worse than when allele frequencies are utilised. If there are several alleles at a locus then the number of genotype categories is large, and a sufficiently large sample of the population would be required to estimate the genotype frequencies. Many categories may not be seen at all, and would be allocated a probability of zero, although in reality that genotype could exist for that particular breed. Of course, the estimation of allele frequencies is subject to the same problem, although to a lesser degree since there are fewer classes to be observed. In practice, it may be better to take either an "ad hoc" approach, and allocate a small probability (say $1/2n$ where n is the number of individual animals sampled) to alleles that appear to have a zero frequency, or to smooth the estimated probabilities (see Krzanowski and Marriott 1995 p17).

Allocation of individual animals to breeds using the method described in this chapter depends on the estimation of multilocus genotype probabilities. These probabilities are subject to sampling fluctuation, since the allele frequencies used to calculate them are estimated from samples drawn from the different breeds. When reporting the results of a test on an individual animal it would be desirable to be able to attach some level of confidence to the probability of its genotype belonging to particular breeds. Chakraborty et al (1993) have described how standard errors and confidence intervals for multilocus genotype probabilities can be evaluated, and give simple approximations to the sampling variance. An alternative approach might be to use bootstrapping to estimate the confidence interval, by resampling the observations on which the allele frequency estimates are based. A new multilocus genotype

probability would be generated for each bootstrap replicate, and 95% confidence intervals could be constructed around the probability of that genotype occurring in each breed.

A short study on the feasibility of inferring ethnic origin of humans from DNA profiles (Evett and Pinchin 1992) concluded that it was possible to provide information on ethnic origin, but that there were limitations. A similar caution might apply to the results presented in this chapter. The problem to be studied must be clearly defined, it is desirable that other information than just genetic information is incorporated, for example in defining a limited set of breeds for which a sample should be tested. It may also be desirable to have an "unknown" category, so that if an animal falls below a certain threshold probability of belonging to the breeds of interest it is classed as "unknown". Even given such limitations genetic markers can, nevertheless, be substantially useful in determining the breed of an animal.

Chapter 7

General discussion

7.1 Genetic characterization of breeds

Preservation of genetic variation in livestock species is necessary if breeds are to be continued to be improved, and if livestock are to be available that will be productive in different environments. Changes in the environment, for example climate change or new disease challenges, will require a range of genotypes to be available in order for these new environments to be exploited. Since the economic cost of conserving all breeds in all species is prohibitive, only a few breeds can be selected for conservation on a global scale (Food and Agriculture Organization 1996). One criterion that might be used to select breeds for conservation would be the possession of unique genetic characteristics. However, it is currently not possible to identify all the genes that contribute to the phenotypic differences between breeds. It has been suggested that breeds that have unique evolutionary histories are those which should be selected for conservation, as they are most likely to carry unique gene combinations not present in other breeds (Hall and Bradley 1995). The evolutionary history of breeds can be inferred by sampling markers from the genome and estimating a general measure of variation between breeds, such as genetic distance, from which phylogenetic relationships among breeds can be constructed. In order to be able to identify breeds that should be conserved, it will be necessary to first survey all existing breeds to establish genetic relationships among them (Food and Agriculture Organization 1996).

In many countries there is already a considerable amount of genetic data available on cattle breeds, that have been collected by national cattle blood typing services over the last thirty or more years. Despite the availability of this information, no studies have been published of a complete analysis of the data collected by a national blood typing laboratory. In this thesis a survey of relationships among thirty-seven European cattle breeds, based on blood type records for almost 19,000 cattle

collected by the Roslin Cattle Blood Typing Service, has been presented. Estimates of the allele frequencies at seven red cell antigen systems (A, B, C, F, S, L, Z) and two serum protein loci (transferrin and albumin) were made, and used to calculate genetic distances between breeds.

The estimation of genetic distances based on blood type gene frequencies was complicated by the fact that many blood type loci are dominant, and in addition the antigenic factors (loci) within the complex blood type systems (B, C and S) are inherited as "phenogroups" or haplotypes. In general, measures of genetic distance assume that there is linkage equilibrium among loci (Nei 1972; Reynolds et al 1983), i.e. the gene frequencies at loci are independent. One method of accounting for this linkage among loci would have been to estimate haplotype frequencies at each of the B, C and S systems. However, the B system in particular is very complex, 18 antigenic factors from the B system were included in this study. The number of haplotypes that this could give rise to is 2^{18}, and there are many more possible phenotypes. The approach taken in previous studies, based on cattle blood types, was to simplify the system by either estimating the frequencies of a small number of common haplotypes and pooling all others into one class, or analysing the component antigenic factors as independent loci.

Two methods of analysing the cattle B blood type system were compared in this thesis; firstly, counting a limited number of factors which were found to be almost independent, a method similar to the estimation of haplotype frequencies and called here the "reduced haplotype" method (Grosclaude et al 1990). The second method analysed antigenic factors as though they were all independent loci, and was termed the "independent loci" method. The "independent loci" method was found to result in genetic distances that were highly positively correlated with genetic distances based on haplotype frequencies. In contrast, the "reduced haplotype" method, although positively correlated with the "independent loci" method, was found to discard too much information due to the grouping of phenotypes. It was particularly uninformative for distinguishing among British breeds.

Ideally, the covariances or linkage disequilibrium between linked loci should be incorporated into the estimates of genetic distance so that redundancy of information can be accounted for (Manly 1986; Dillmann et al 1997). Estimates of the pairwise linkage disequilibria between loci of the B system in nine different breeds indicated that patterns of linkage disequilibrium were breed specific. This made it difficult for the covariances (disequilibria) between loci to be incorporated into the measure of genetic distance, as the methods currently available assume that the covariances between loci are the same for all populations. However, since the covariances were also found to be generally small it was concluded that treating the linked loci of the complex blood type systems as independent was acceptable when estimating genetic distance between breeds.

Linkage among loci was also a factor that had to be taken into account when estimating the sampling variance of genetic distance. The use of bootstrapping as a method for estimating sampling variances of genetic distance was investigated. The sampling variance of genetic distance can be partitioned into two components, the between or interlocus variance, and the within or intralocus variance. Bootstrapping by resampling individual genotypes (or phenotypes in the case of blood types) allowed the intralocus variance to be estimated, while resampling loci allowed the interlocus variance to be estimated. Using bootstrapping it was also possible to account for linkage between loci when estimating the sampling variance, by sampling genotypes at linked loci together. A comparison of two bootstrap sampling strategies, firstly sampling linked loci together and secondly sampling all loci, even linked ones, independently, indicated that there was little difference in the two estimates of the variance. This confirmed that the correlations between the gene frequencies at loci of the B system are weak, since it was expected that ignoring the correlations between linked loci would result in the variance being underestimated.

The survey of thirty-seven European cattle breeds revealed two major groupings of breeds; with breeds from France, Italy, Germany, Switzerland and the Channel Islands forming one group and breeds from Britain and Northern Europe the other group. The breeds tended to form groups that were based on their geographic origins

rather than breed purpose, i.e. beef and dairy breeds did not form separate groups. This survey, however, was based on a limited number of blood type and serum protein loci, which are located on just nine chromosomal regions. The reliability of estimated breed relationships is therefore low, and this was confirmed by the small bootstrap values obtained for the nodes of the phylogenetic trees. Drift may also have had a major influence on the apparent genetic uniqueness of some breeds, as the genetic distance between populations is accelerated by population bottlenecks (Takezaki and Nei 1996). It was notable that a number of breeds, e.g. British White, White Park, Lincoln Red, Sussex, that were found to be distant from all the other breeds, have small population sizes which makes it likely that they have undergone more genetic drift than other breeds. These breeds were also observed to have low heterozygosities and a small number of alleles, which lend weight to the supposition that genetic drift has played an important role in their histories. In general, breeds from mainland Britain were found to be less heterozygous than continental breeds such as the Limousin, Charolais and Blonde d'Aquitaine.

7.2 Genetic variation within breeds

Within breeds the population is structured, possibly into herd book divisions defined by some phenotypic criterion (e.g. polled and horned), and certainly into herds and families. For many breeds there are populations in several countries, with each population being derived from separate founder groups and selected for different objectives. An analysis of the genetic structure of one such breed, the Hereford, was presented in this thesis. Genetic differences between Hereford populations in the United Kingdom, Ireland, Canada, Sweden and New Zealand were studied using blood type data that had been collected over thirty years. The British and Canadian populations were divided by polled or horned status, ancestry (whether of mixed British/Canadian ancestry or not) and year of birth. The effect of interbreeding with the Canadian population on the genetic structure of the British Hereford population was considered.

All the Hereford populations were found to group together and were distinct from other breeds. Horned and polled populations did not group together, but groups from the same geographic origin did, for example the British populations clustered and so did the Canadian populations. There was also a significant structuring of the Hereford population into herds. Most of the herds studied were found to be significantly different from one another, which suggested limited gene flow between them. This might be partly accounted for by the fact that the herds were of either polled or horned animals, and there tends to be a clear definition between the two, i.e. some herds consist only of polled animals others of only horned animals. There was also variation in the breeding strategies employed, some herds consisting only of animals with 100% British ancestry and others consisting of Canadian or part-Canadian animals.

The Canadian groups (both polled and horned) were found to be the least heterozygous of all the Hereford groups. Heterozygosity within all sections of the British population was seen to have declined over time. It was most marked in the group with pure British ancestry, probably a consequence of the severe decline in numbers of animals in this category, from several thousand to just 400. The majority of Herefords in the UK now have mixed British and Canadian ('hybrid') ancestry. This 'hybrid' group was observed to have the highest heterozygosity, as might be expected of a cross between two previously isolated populations. However, even within this group heterozygosity was seen to be declining with time, perhaps in part due to the increasing proportion of genes from the least variable of the parental populations. The proportion of Canadian genes in the 'hybrid' population was estimated at 0.65 ($\pm$ 0.21). The entire replacement of the British Hereford population with Canadian-bred Herefords may result in an accelerated decline in genetic variation. Genetic improvement, whether by within population selection or by introgression from other populations, should be balanced against the maintenance of genetic diversity, in order to protect the long-term future of the breed.

Introgression and the replacement of native populations with animals from other countries have already had a major impact on dairy breeds. The replacement of other

black-and-white cattle strains by the North American Holstein has reduced the genetic diversity of the global population of black-and-whites (Goddard 1992). This loss of variation is not due to the introduction of foreign bloodlines *per se*, but due to a reduction in the effective population size of the breed caused by the widespread use of a small number of selected sires (Goddard 1992). The reduction in effective population sizes of breeds has been accelerated by the use of reproductive technologies such as AI and embryo transfer. Although introducing Canadian and North American bloodlines can vastly improve performance, there may be negative effects on traits such as fertility (Lidauer and Mantysaari 1996). Inbreeding depression has also been shown to directly affect the production traits breeders are desirous of improving, for example milk and protein yield in dairy cattle are reduced in inbred animals (Miglior et al 1992, 1995b; Sigurdsson and Jonmundsson 1995). Inbreeding may also have an impact on disease resistance (Miglior et al 1995a), and increases the risk of inherited disease. For example, BLAD (bovine leukocyte adhesion deficiency) (Mirck et al 1995) and DUMPS (uridine monophosphate synthase deficiency) (Schwenger et al 1993) in the Holstein-Friesian are both inherited defects that have risen to high frequencies in the population, and can be traced back to a single founder.

Goddard (1992) has suggested that a better global breeding programme for the Holstein-Friesian would take into account inbreeding depression when sires are selected. Meuwissen and Woolliams (1994) have shown that the effective population size of livestock populations required to prevent a decline in fitness (fitness is defined as the number of offspring an individual has compared with the average), due to inbreeding depression, is larger than that required to maximize genetic gain. This implies that unless breeding programmes are specifically designed to take account of inbreeding depression, then livestock effective population sizes are likely to become too small to prevent a decline in fitness.

7.3 Genetic identification of breeds

Crossing of breeds with other more productive breeds is frequently carried out in order to upgrade the performance of the original breed. In some rare breeds outcrossing has occurred because a severe reduction in population size has necessitated the introduction of animals from another breed. However, a basic principle of breed conservation is the promotion of pure-breeding (Hall and Bradley 1995). This is achieved, in most countries, by keeping pedigree records and registering pedigree animals. In recent years, the accuracy of pedigree records has been checked by blood or DNA typing. There has also been interest in the use of genetic markers to identify the breed of an animal, without recourse to pedigree information. A test of breed identity would be useful in situations where pedigree records do not exist, and for validating the origin of livestock products that have been marketed under the breed "brand" name. In chapter six two types of genetic marker were compared for their efficiency in distinguishing among cattle breeds, and the number of markers required to distinguish between purebred European breeds and some of their crosses was investigated.

Microsatellite markers were found to be more powerful than diallelic markers for distinguishing among breeds. On average, six diallelic markers were required to achieve the same power of discrimination as a single microsatellite marker. The efficiency of discrimination among breeds could be improved by selecting markers on their individual error rates. By choosing markers that were the most discriminatory the number of microsatellite markers required to distinguish among seven European breeds could be reduced from about twelve to six. More markers were required to distinguish among pure and crossbred animals. In the example given in chapter six, where discrimination was among five pure breeds and four Friesian crosses, at least thirty randomly selected markers were required to achieve error rates of less than 5%. If the most discriminatory markers were chosen then it was predicted the same error rates could be achieved with about fifteen markers.

The most discriminatory markers were those with high average heterozygosities, and greater numbers of observed alleles. On closer examination, the best three markers for discriminating among pure and crossbreed European breeds were found to have diagnostic alleles in some breeds. At the marker ETH131 nineteen from a possible twenty-two alleles were observed in the European breeds, with three alleles diagnostic for the Simmental and one allele diagnostic for the Holstein-Friesian. A further nine alleles were only present in two, three or four of the seven breeds. The marker BTMICROS was found to have a diagnostic allele for the Hereford (169 bp), at a reasonably high frequency of 0.47, and also an allele present only in the Holstein-Friesian and Jersey (159 bp). BM2113 had a diagnostic allele for the Jersey (131 bp) and another allele only present in the Charolais and Simmental (133 bp).

The selection of markers to be used in breed identification will depend on the problem to be addressed. For any particular set of breeds a set of markers could be identified which provide optimal power of discrimination. Ideally each breed would be identified with a set of markers that have alleles specific to that breed. This would enable breed identification to be carried out using just one or two markers for every breed to be compared, or less if markers were available that had alleles specific to a range of different breeds. Given the range of markers currently available in livestock, the use of DNA markers to provide a test of breed identity is now a real possibility. There are, however, practical limitations to the test. For example, the more closely related populations or breeds are, the more markers are required to distinguish between them. It would not currently be feasible to provide a DNA test for discriminating between, for example, British and Canadian populations of the Hereford as it would require hundreds of markers. The cost of the test would be likely to be more than the customer would want to pay.

7.4 Future work

The survey of breed relationships described in this thesis was based on nine chromosomal regions, coding for the cattle red cell antigens and two serum proteins. The true genetic history of a breed is made up of a fine web of individual gene

phylogenies (Maddison 1996). Different parts of the genome have different histories and trees constructed from a few loci may not reflect the true history of the organism (Avise 1989). The full picture can only be obtained by using a variety of different markers and by sampling a sufficient proportion of the genome. Maddison (1996) described phylogenies in the following way: "A phylogeny is like a gene-history cloud with the greatest density of gene histories forming a tree-like shape, but with a potentially broad cloud of other histories completing the picture ..... phylogenetic history is composed simultaneously of different relationships for different portions of the genome. When we see a phylogenetic tree of simple lines we must realize it tells only part of the story - something like the modal gene history, and not the variance about it".

Microsatellite loci are generally considered to be the most powerful genetic markers for studying population differentiation and making phylogenetic inferences about very close related groups (Goldstein et al 1995). The availability of microsatellite marker maps for most livestock species including cattle, pigs, sheep and chickens, generated as part of the genome mapping projects, also makes the use of these markers appealing. Measures of genetic distance and methods of phylogenetic inference based on explicit models of microsatellite evolution have recently been developed (reviewed in Goldstein and Pollock 1997), providing more accurate means of identifying relationships among populations. Practical difficulties with the use of microsatellite markers, that are yet to be resolved, include the difficulty of determining fragment lengths of alleles which complicates automation of the procedure. Related to this is the problem of comparing fragment lengths across laboratories, and the standardization of results. In addition, because of the difficulty in determining allele lengths, there is a high risk of genotyping error adding considerable "noise" to the results. The genotyping error can only be detected when family information is available and in most diversity studies this information is not available, as individuals sampled are selected for unrelatedness. Finally, the recommended number of loci for diversity studies is at least fifty (Takezaki and Nei 1996), and the cost and effort of typing large numbers of microsatellite loci may well be prohibitive.

New technologies for genome analysis are constantly being developed, and the future of genetic diversity studies may lie with other types of marker (Goldstein and Pollock 1997). Technologies that will allow large numbers of loci to be economically scanned are already available e.g. AFLPs, SNPs. Some have already been applied to the problem of assessing diversity, particularly in plant populations (Lu et al 1996; Paul et al 1997; Hongtrakul et al 1997), and also in livestock populations (Ajmone-Marsan et al 1996). These types of marker will allow large areas of the genome to be scanned, and small genetic differences between breeds to be detected. They may also have a role in mapping functional genes that differ among breeds.

Another approach to phylogenetic analysis that would provide valuable insights into population history, particularly for intraspecific populations, would be to study haplotypes in specific genome regions and the cladistic relationships among these haplotypes (Templeton et al 1987; Crandall and Templeton 1993). The method developed by Templeton et al (1987) also allows phenotypically important mutations to be identified, when the genome region studied contains a gene that affects a trait of interest. For example, Lagziel et al (1996) found associations between haplotypes at the bovine growth hormone gene and milk protein percentage, and were able to distinguish effects associated with haplotypes originating from either *Bos taurus* or *Bos indicus*. If several genome regions are studied then the resulting cladograms would not only give a more detailed picture of breed development and history, but would also help to elucidate how genome differences between breeds influence their phenotypic differences. The analysis of haplotypes, and linkage disequilibria between the component loci of the haplotype, can also provide detailed information on the history of populations and may be particularly useful for dating population divergence (Tishkoff et al 1996).

Understanding the genetic differences between breeds that give rise to variation in phenotypic characteristics will eventually be accomplished as progress is made in livestock genome mapping. Several hundred markers have now been placed on the

bovine genome map (Barendse et al 1997), which together with information on the phenotypes of animals will permit the genes underlying traits to be mapped. Eventually the fine-mapping of quantitative trait loci (QTL) affecting important economic traits, such as milk production, will make possible fine-scale comparisons between breeds (Georges and Andersson 1996).

It will also become possible to ask more complex questions about the evolution and history of different breeds. For example, there is evidence in cereal crop species that different species can share the same QTL, particularly for traits that have proved important in the domestication of the species (Paterson et al 1995). However, it is not known whether the correspondence in QTL found in interspecific crosses is also found in more closely related genotypes, such as elite crop cultivars (Beavis et al 1991). It has been suggested that crop gene pools might be homogeneous at mutant alleles with large effects, these alleles being fixed during the domestication process, and that variation within the gene pools are a result of subsequent mutation at a large number of other loci with smaller effects (Paterson 1995). If this is also the case in domestic animal populations, then different breeds might be expected to share QTL alleles of large effect (this will also be the case if these alleles have been introgressed from one breed to another). Differences between breeds might then be the result of variation at other modifier loci. There is already some information about breed differences being revealed by the mapping of major genes. Mapping of the double-muscling locus (Grobet et al 1997), and comparison of the gene in five different breeds revealed allelic heterogeneity. Two of the three breeds with the double-muscled phenotype (Belgian Blue and Asturiana) shared the same deletion, but the third (Maine-Anjou) had an apparently normal genotype suggesting that another mutation (allele) was responsible for its phenotype.

In summary, modern molecular tools are providing new opportunities to understand, at the genetic level, what defines a breed and how the essential characteristics of our livestock breeds might be conserved for future generations.

7.5 Conclusions

The main conclusions to be drawn from this thesis are:

1. Analysis of the structure of the Hereford breed using genetic markers showed that some populations of Herefords are less variable others, and suggested that substitution of one national population by another may result in a faster decline in genetic variation within the breed. Future breeding programmes for livestock may need to include strategies that balance genetic improvement with the maintenance of genetic variation. This will be important not just for rare breeds or small populations but even for more numerous breeds, such as the Hereford and the Holstein-Friesian, where national populations of the breed have, as a result of exportation of animals and semen for breeding, become one global population. Potential strategies for achieving a balance between selection and inbreeding or loss of variation, based on pedigree information, have been proposed by Goddard (1992), Meuwissen and Woolliams (1994) and Villanueva and Woolliams (1997). Further work is required, both to investigate the use of such breeding schemes in rare breeds of livestock, and to investigate whether the integration of marker information into the scheme might be advantageous.

2. Genetic markers will be useful for providing a test of breed identity, but with certain limitations. The question that is being asked must be clearly defined, at least the methods presented in this thesis do not allow for open-ended questions such as "is this breed A or something else?" While it is possible to distinguish among pure breeds with a relatively small number of markers, discriminating among closely related populations, for example crossbreeds or different populations of the same breed, requires a larger number of markers. The cost of carrying out this type of test, other than for pure breeds, may not yet be affordable.

3. The analysis of European cattle breeds, based on genetic distances, that is presented in this thesis, while giving an overall view of the genetic differences among breeds, cannot answer specific questions about how breeds have evolved or how different gene variants may affect a breed's performance. However, new developments in genome analysis, and the existence of genome maps for a number of livestock species, can provide the tools for more detailed analyses. In the future, it is likely that more complex questions concerning the evolution of breeds and the genetic differences between them can be answered.

Appendix I

Blood typing reagents (antibodies) used by the Roslin Cattle Blood Typing Service between 1966 and 1997

Date	System / Lab Nos	A1	A2	D	H	Z'	B1	B2	B3	G1	G2	G3	I1	I2	K1	K2	O1	O2	O3	O4	O5	P1	P2	Q1	Q2	T1	T2	Y1	Y2	A'1	A'2	B'	D'	E'1	E'2	E'3	E'4	F'1	F'2
25/07/66	A		X					X			X			X		X	X						X	X	X				X	X			X			X			
25/10/66	A		X					X			X			X		X	X						X	X	X		X		X	X			X			X			
26/01/67	B		X	X		X		X			X		X	X		X	X						X	X	X		X		X	X			X			X			
26/04/67	B		X					X			X			X		X	X						X	X	X		X		X	X			X			X			
26/07/67	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
26/10/67	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
25/01/68	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
30/04/68	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
30/07/68	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
31/10/68	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
30/01/69	B		X					X			X		X	X		X	X						X	X	X		X		X	X			X			X			
27/03/69	B		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
25/06/69	B		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
04/08/69	B		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
05/11/69	C		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
05/02/70	C		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
05/05/70	C		X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
05/08/70	C	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X			X			
04/11/70	C	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
04/03/71	C	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
10/06/71	C	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
27/07/71	D	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
28/10/71	D	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
28/02/72	D	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X			X		
29/06/72	D	X	X					X			X		X	X		X	X	X	X				X	X	X		X		X	X			X	X		X			
26/10/72	D	X	X					X			X		X	X		X	X	X	X				X		X		X		X	X			X	X			X		
28/02/73	D	X	X					X			X		X	X		X	X	X	X				X		X		X		X	X			X		X				
18/04/73	E	X	X					X			X		X	X		X	X		X				X		X		X		X	X			X		X				
21/08/73	E	X	X					X			X		X	X		X	X		X				X		X		X		X	X			X						
20/12/73	E	X	X					X			X		X	X		X	X		X				X		X		X		X	X			X		X				
23/04/74	E	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
13/05/74	E	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
14/08/74	F	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
12/12/74	F	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
15/04/75	F	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
15/05/75	G	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
19/08/75	G	X	X					X			X		X	X			X		X				X		X		X		X	X			X		X				
17/12/75	G	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
18/03/76	G	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
04/05/76	G	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
05/08/76	G	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
04/11/76	H	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
05/03/77	H	X	X					X			X		X	X					X				X		X		X		X	X			X		X				

Date	System / Lab Nos	A1	A2	D	H	Z'	B1	B2	B3	G1	G2	G3	I1	I2	K1	K2	O1	O2	O3	O4	O5	P1	P2	Q1	Q2	T1	T2	Y1	Y2	A'1	A'2	B'	D'	E'1	E'2	E'3	E'4	F'1	F'2
20/04/77	H	X	X					X			X		X	X					X				X		X		X		X	X			X		X	X			
20/07/77	K	X	X					X			X		X	X					X				X		X		X		X	X			X		X	X			
20/10/77	K	X	X					X			X		X	X					X				X		X		X		X	X			X		X	X			
24/01/78	K	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
20/04/78	K	X	X					X			X		X	X					X				X		X		X		X	X			X		X				
10/05/78	K	X	X					X			X		X	X					X				X		X		X		X	X			X		X	X			
10/08/78	L	X	X					X			X		X	X		X			X				X		X		X		X	X			X		X	X			
08/11/78	L	X						X			X		X	X					X				X		X		X		X	X			X		X	X			
12/02/79	L	X						X			X		X	X					X				X		X				X	X			X		X	X			
10/05/79	L	X						X			X		X	X					X				X		X		X		X	X			X		X	X			
12/06/79	L		X					X			X		X	X					X				X		X		X		X	X			X		X	X			
15/10/79	L	X						X			X		X	X					X				X	X	X		X		X	X			X		X	X			
14/02/80	L	X						X			X		X	X		X			X				X	X	X		X		X	X	X		X		X				
24/04/80	M	X									X		X	X		X			X				X	X	X		X		X	X	X		X		X				
27/08/80	M		X								X		X	X		X			X				X	X	X		X		X	X	X		X		X				
23/12/80	M		X			X	X	X			X		X	X		X					X		X	X	X		X	X	X	X	X		X		X				
09/02/81	M		X			X		X			X		X	X					X				X	X	X		X	X	X	X	X		X		X				
09/06/81	M		X			X		X			X		X	X									X	X	X		X	X	X	X	X		X		X				
07/10/81	M		X			X		X			X		X	X	X	X			X	X	X		X	X	X			X	X	X	X		X		X				X
10/12/81	N		X			X		X			X		X	X	X	X			X	X			X	X	X		X	X	X	X	X		X		X				X
06/04/82	N		X			X		X			X		X	X	X	X				X			X	X	X		X	X	X	X	X		X		X				X
05/08/82	N		X			X		X			X		X	X	X	X				X			X	X	X		X	X	X	X	X		X						X
15/09/82	N	X				X	X			X			X	X	X	X	X	X		X		X		X	X	X		X	X	X	X		X		X	X			X
15/12/82	N	X				X	X			X			X	X		X	X	X		X		X		X	X	X		X	X	X	X		X		X	X			X
16/03/83	N	X				X	X			X			X	X	X	X	X	X		X		X		X	X	X		X	X	X	X		X		X				X
06/04/83	P	X			X	X	X		X	X		X	X	X	X	X	X	X		X		X		X	X	X		X	X	X	X	X	X		X	X			X
06/07/83	P	X				X	X			X			X	X	X	X	X	X		X		X		X	X	X		X	X	X	X		X		X				X
06/10/83	P	X				X	X			X			X	X	X	X	X	X		X		X		X	X	X		X	X	X	X		X		X				X
16/11/83	P	X				X	X			X			X	X	X	X	X	X		X		X		X	X	X		X	X	X	X		X		X				X
11/04/84	P	X			X	X	X		X	X		X	X	X	X	X	X	X		X		X		X	X	X		X	X	X	X	X	X		X	X			X
13/07/84	P	X	X		X	X	X			X			X	X	X		X	X		X		X		X	X	X		X	X	X	X	X	X		X	X			X
12/10/84	P	X	X		X	X	X			X			X	X	X		X	X		X		X		X	X	X			X	X	X	X	X			X			X
15/01/85	P	X	X		X	X	X			X			X	X	X		X	X		X		X		X	X	X			X	X	X	X	X		X	X			X
13/02/85	Q		X		X	X	X			X			X	X	X		X	X		X		X		X	X	X			X	X	X	X	X			X			X
16/05/85	Q		X			X	X			X			X	X	X		X			X		X		X	X	X			X	X	X	X	X			X			X
15/08/85	Q		X			X	X			X			X	X	X	X	X			X	X			X	X	X			X	X	X	X	X			X			X
15/11/85	Q	X				X	X			X			X	X		X	X			X	X	X	X	X	X	X			X	X	X	X	X			X			X
14/02/86	Q	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X		X	X			X			X
15/05/86	Q	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X		X	X			X			X
15/08/86	R	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X	X	X	X			X			X
19/11/86	R	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X	X	X	X			X			X
17/02/87	R	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X	X		X			X			X
15/05/87	R	X				X	X			X			X	X		X	X			X	X	X		X	X	X			X	X	X		X	X		X			X

Date	System / Lab Nos	A1	A2	D	H	Z'	B1	B2	B3	G1	G2	G3	I1	I2	K1	K2	O1	O2	O3	O4	O5	P1	P2	Q1	Q2	T1	T2	Y1	Y2	A'1	A'2	B'	D'	E'1	E'2	E'3	E'4	F'1	F'2
01/07/87	R	X				X	X			X			X	X			X			X	X	X		X	X	X			X	X	X		X	X		X			X	
21/08/87	R	X				X	X			X			X	X			X			X	X	X		X	X	X			X	X	X		X	X					X	
25/11/87	S	X				X	X			X			X	X	X		X			X	X	X		X	X	X			X	X	X		X	X					X	
29/01/88	S	X				X	X			X			X	X	X						X	X		X	X	X			X	X	X			X					X	
25/02/88	S	X					X			X			X	X	X		X	X	X	X		X		X	X	X			X	X	X		X	X					X	
27/05/88	S	X					X			X			X	X	X		X	X		X		X		X	X	X			X	X	X		X	X					X	
18/08/88	S	X					X			X			X	X	X			X		X		X		X	X	X			X	X	X		X	X					X	
24/11/88	S	X					X			X			X	X	X			X		X		X		X	X	X			X	X	X		X	X					X	
17/02/89	T	X					X			X			X	X	X					X		X		X	X	X			X	X	X		X	X					X	
25/05/89	T	X					X			X			X	X	X							X		X	X	X			X	X	X		X	X					X	
25/08/89	T	X					X			X			X	X	X			X				X		X	X	X			X	X			X	X					X	
24/11/89	T	X					X			X			X	X	X		X					X			X	X			X	X			X	X					X	
23/02/90	V	X					X			X			X	X	X							X			X	X			X	X			X	X					X	
25/05/90	V	X					X			X			X	X	X							X			X	X			X	X			X	X					X	
23/08/90	V	X					X			X			X	X	X							X			X	X			X	X			X		X				X	
24/11/90	V	X					X			X			X	X	X							X			X	X			X	X			X		X				X	
22/02/91	V	X	X				X			X			X	X	X							X			X	X			X	X			X		X				X	
24/05/91	V	X	X				X			X			X	X	X							X			X	X			X	X			X		X				X	
23/08/91	V	X	X				X			X			X	X	X							X			X	X			X	X			X		X				X	
22/11/91	V	X	X					X			X		X	X	X							X			X		X		X	X		X	X		X			X	X
21/02/92	V	X	X				X			X			X	X	X		X	X				X			X	X			X	X		X	X		X			X	X
22/05/92	V	X	X				X	X		X			X	X	X		X	X				X			X		X		X	X	X	X	X		X			X	X
21/08/92	V	X	X				X	X		X			X	X	X		X	X				X			X		X		X	X	X	X	X		X			X	X
20/11/92	V19	X	X				X	X		X			X	X			X	X				X			X		X	X	X	X	X	X	X		X			X	X
24/02/93	V20	X	X				X	X		X			X	X			X	X				X			X		X	X	X	X	X	X	X		X			X	X
26/05/93	V21	X	X				X	X		X			X	X			X	X				X					X	X	X	X	X	X	X		X			X	X
25/08/93	V23	X	X				X	X		X			X	X			X	X				X			X		X	X	X	X	X	X	X		X	X		X	X
26/11/93	V25	X	X				X	X		X			X	X				X				X			X		X	X	X	X	X	X	X		X	X		X	X
25/02/94	V27	X	X				X	X		X			X	X				X				X			X		X	X	X	X	X	X	X		X	X		X	X
25/05/94	V28	X	X					X			X		X	X			X						X		X		X	X	X	X			X		X		X		
24/08/94	V30	X	X				X			X			X	X			X					X			X			X	X	X			X		X		X		
23/11/94	V32	X	X				X			X			X	X	X		X					X		X			X	X	X	X			X		X		X		
22/02/95	V34	X	X				X			X			X	X	X		X					X		X			X	X	X	X			X		X		X		
24/05/95	V36	X	X				X	X		X			X	X		X	X		X			X		X		X		X	X	X		X	X		X		X		
25/08/95	V38	X	X				X				X		X	X		X	X		X			X			X		X	X	X	X		X	X		X		X		
24/11/95	V40		X				X				X		X	X	X		X				X		X		X	X	X	X	X	X		X	X		X		X		
21/02/96	V42		X				X				X		X	X	X		X				X		X		X	X	X	X	X	X		X	X		X		X		
29/05/96	V44	X						X			X		X		X		X				X	X			X	X		X	X	X		X	X		X		X		
28/08/96	V46	X				X		X			X		X		X		X	X				X		X		X			X	X		X	X		X		X		
20/11/96	V49	X						X			X		X		X		X	X				X			X	X			X	X		X	X		X		X		
05/03/97	V51	X				X		X			X		X		X		X	X				X			X	X			X	X		X	X		X		X		

Date	System Lab Nos	G'1	G'2	I'1	I'2	J'1	J'2	K'	O'1	O'2	P'1	P'2	Q'	Y'	A"	B"1	B"2	D"	F"	G"1	G"2	I"	J"	O"	C C1	C2	E	R1	R2	W1	W2	X1	X2	X3	X'	C'	L'	C"1	C"2
25/07/66	A	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
25/10/66	A	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
26/01/67	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
26/04/67	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
26/07/67	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
26/10/67	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
25/01/68	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
30/04/68	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
30/07/68	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
31/10/68	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
30/01/69	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
27/03/69	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
25/06/69	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
04/08/69	B	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
05/11/69	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
05/02/70	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
05/05/70	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
05/08/70	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
04/11/70	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
04/03/71	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
10/06/71	C	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
27/07/71	D	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
28/10/71	D	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
28/02/72	D	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
29/06/72	D	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
26/10/72	D	X		X		X		X	X		X		X	X												X		X	X		X	X	X					
28/02/73	D	X		X		X		X	X		X		X	X												X		X	X	X	X	X	X					
18/04/73	E	X		X		X		X	X		X		X	X												X		X			X	X	X					
21/08/73	E	X		X		X		X	X		X		X	X												X		X			X	X	X					
20/12/73	E	X		X		X		X	X		X		X	X												X		X			X	X	X					
23/04/74	E			X		X		X			X		X	X												X		X			X	X	X					
13/05/74	E	X		X		X		X	X		X		X	X												X		X			X	X	X					
14/08/74	F			X		X		X	X		X		X	X												X		X			X	X	X					
12/12/74	F	X		X		X		X	X		X		X	X												X		X			X	X	X					
15/04/75	F	X		X		X		X	X		X		X	X												X		X			X	X	X					
15/05/75	G			X		X		X	X		X		X	X												X		X			X	X	X					
19/08/75	G							X	X		X		X	X												X		X			X	X	X					
17/12/75	G	X		X		X		X	X		X		X													X		X			X	X	X					
18/03/76	G	X		X				X	X		X		X													X		X			X	X	X					
04/05/76	G	X		X				X	X		X		X													X		X			X	X	X					
05/08/76	G	X		X				X	X		X		X													X		X			X	X	X					
04/11/76	H	X		X				X	X		X		X													X		X			X	X	X					
05/03/77	H	X		X				X	X		X		X													X		X			X	X	X					

Date	System Lab Nos	G'1	G'2	I'1	I'2	J'1	J'2	K'	O'1	O'2	P'1	P'2	Q'	Y'	A"	B"1	B"2	D"	F"	G"1	G"2	I"	J"	O"	C C1	C2	E	R1	R2	W1	W2	X1	X2	X3	X'	C'	L'	C"1	C"2
20/04/77	H	X		X				X	X		X		X												X	X		X			X	X	X					
20/07/77	K	X		X				X	X		X		X												X	X		X			X	X	X					
20/10/77	K	X		X				X	X		X		X												X	X		X			X	X	X					
24/01/78	K	X		X				X	X		X		X												X	X		X			X	X	X					
20/04/78	K	X		X				X	X		X		X												X	X		X			X	X	X					
10/05/78	K	X		X				X	X		X		X													X		X			X	X	X					
10/08/78	L	X		X				X	X		X		X												X	X		X			X	X	X					
08/11/78	L	X		X				X	X		X		X												X	X		X			X	X	X					
12/02/79	L	X		X				X	X		X		X												X	X		X			X	X	X					
10/05/79	L			X				X	X		X		X												X	X		X			X	X	X					
12/06/79	L			X				X	X		X		X													X		X			X	X						
15/10/79	L			X				X	X		X		X												X	X		X			X	X	X					
14/02/80	L	X		X				X	X		X		X	X											X	X		X			X	X	X			X		
24/04/80	M	X		X				X	X		X		X												X	X		X			X	X	X					
27/08/80	M	X		X				X	X		X		X	X											X	X		X			X	X	X					
23/12/80	M	X		X		X		X	X		X		X	X											X	X		X			X	X	X					
09/02/81	M	X		X				X	X		X		X	X											X	X		X			X	X	X			X		
09/06/81	M	X		X				X	X		X		X	X											X	X		X			X	X	X			X		
07/10/81	M	X		X	X	X		X	X		X		X	X											X	X		X			X	X	X			X		
10/12/81	N	X		X	X	X					X		X	X											X	X		X			X	X	X			X		
06/04/82	N	X		X	X	X		X	X	X	X		X	X											X	X		X			X	X	X			X		
05/08/82	N	X		X	X	X		X	X		X		X	X											X	X		X			X	X	X			X		
15/09/82	N	X		X	X	X		X	X		X		X	X						X					X	X		X		X		X	X			X		
15/12/82	N	X		X	X	X		X	X		X		X	X						X					X	X		X		X		X	X			X		
16/03/83	N	X		X	X	X		X	X		X		X	X											X	X		X		X		X	X			X		
06/04/83	P	X	X	X		X		X	X		X		X	X		X									X	X		X		X		X	X					
06/07/83	P	X		X		X		X	X		X		X	X											X	X		X		X		X	X			X		
06/10/83	P	X	X	X	X	X		X	X		X		X	X											X	X		X		X		X	X			X		
16/11/83	P	X	X	X	X	X		X	X		X		X	X											X	X		X		X		X	X			X		
11/04/84	P	X	X	X		X		X	X		X		X	X		X									X	X		X		X		X	X					
13/07/84	P	X	X	X		X		X	X		X		X	X		X									X	X		X		X		X	X			X		
12/10/84	P	X	X	X		X		X	X		X		X	X		X									X			X		X		X	X					
15/01/85	P	X	X	X		X		X	X		X		X	X		X									X	X		X		X		X	X			X		
13/02/85	Q	X	X	X		X		X	X		X		X	X		X									X	X		X		X		X	X			X		
16/05/85	Q	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
15/08/85	Q	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
15/11/85	Q	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
14/02/86	Q	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
15/05/86	Q	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
15/08/86	R	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
19/11/86	R	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
17/02/87	R	X	X	X		X		X	X		X		X	X		X	X								X	X		X		X		X	X			X		
15/05/87	R	X	X	X		X		X			X		X	X			X								X	X		X		X		X	X			X		

Date	System Lab Nos	G'1	G'2	I'1	I'2	J'1	J'2	K'	O'1	O'2	P'1	P'2	Q'	Y'	A"	B"1	B"2	D"	F"	G"1	G"2	I"	J"	O"	C C1	C2	E	R1	R2	W1	W2	X1	X2	X3	X'	C'	L'	C"1	C"2
01/07/87	R	X	X	X		X		X			X		X	X			X								X	X		X		X		X	X				X		
21/08/87	R	X	X	X		X		X			X		X	X			X								X	X		X		X		X	X				X		
25/11/87	S	X	X			X		X	X		X		X	X			X								X	X		X		X		X	X				X		
29/01/88	S	X	X	X		X		X			X		X	X			X								X	X		X		X		X	X				X		
25/02/88	S	X	X	X		X		X			X		X	X			X								X	X		X		X		X	X				X		
27/05/88	S	X	X	X		X		X			X		X	X											X	X		X		X		X	X				X		
18/08/88	S	X	X			X		X			X		X	X			X								X	X		X		X		X	X				X		
24/11/88	S	X	X			X		X			X		X	X											X	X		X		X		X	X				X		
17/02/89	T	X	X			X		X	X		X		X	X											X	X		X		X		X	X				X		
25/05/89	T	X	X			X		X	X		X		X	X											X	X		X		X		X	X				X		
25/08/89	T	X				X		X	X		X		X	X											X	X		X		X		X	X				X		
24/11/89	T	X				X		X	X		X			X											X	X		X		X		X	X						
23/02/90	V	X				X		X	X		X			X											X	X		X		X		X	X				X		
25/05/90	V	X				X		X	X		X			X											X	X		X		X		X	X				X		
23/08/90	V	X		X		X		X	X		X			X											X	X		X		X		X	X						
24/11/90	V	X		X		X		X	X		X			X											X	X		X		X		X	X						
22/02/91	V	X		X		X		X	X		X			X											X	X		X		X		X							
24/05/91	V	X		X		X		X	X		X		X	X											X	X		X		X		X							
23/08/91	V	X		X		X		X	X		X		X	X											X	X		X		X		X							
22/11/91	V	X		X		X		X	X		X			X											X	X		X		X		X							
21/02/92	V	X		X		X		X	X		X			X		X									X	X		X		X		X							
22/05/92	V	X		X		X		X	X		X			X		X	X								X	X		X		X		X							
21/08/92	V	X		X		X		X	X		X			X		X									X	X		X		X		X							
20/11/92	V19	X		X		X		X	X		X			X		X									X	X		X		X		X	X						
24/02/93	V20	X		X		X		X	X		X			X		X									X	X		X		X		X	X						
26/05/93	V21	X		X		X		X	X		X			X		X									X	X		X		X		X	X						
25/08/93	V23	X		X		X		X	X		X			X		X									X	X		X		X		X	X						
26/11/93	V25	X		X		X		X	X		X		X	X		X									X	X		X		X		X	X						
25/02/94	V27	X		X		X		X	X		X		X	X		X									X	X		X		X		X	X						
25/05/94	V28	X		X		X		X	X		X		X	X		X									X	X		X		X		X	X						
24/08/94	V30	X		X		X		X	X		X		X	X		X						X			X	X		X				X	X						
23/11/94	V32	X		X		X		X	X		X		X	X								X			X	X		X		X		X	X						
22/02/95	V34	X		X		X		X	X		X		X	X								X			X	X		X		X		X	X						
24/05/95	V36	X		X		X		X			X			X								X			X	X		X		X		X	X						
25/08/95	V38	X		X		X		X	X		X			X		X						X			X	X		X		X		X	X						
24/11/95	V40	X		X		X		X	X		X			X								X			X	X		X		X		X	X						
21/02/96	V42	X		X		X		X	X		X			X		X						X				X		X	X	X		X	X						
29/05/96	V44	X		X		X		X			X			X								X			X	X		X	X	X		X	X						
28/08/96	V46	X		X		X		X	X		X			X			X					X				X		X		X		X	X						
20/11/96	V49	X		X		X		X	X		X			X			X					X				X		X		X		X							
05/03/97	V51	X		X		X	X	X	X		X			X			X					X				X		X		X		X							

Date	System Lab Nos	F F1	F2	V1	V2	N'1	N'2	V'	J J1	J2	L L	M M1	M2	M'	S S1	S2	U1	U2	H'	U'1	U'2	H"	S"	U"	Z Z	R' R'	S'1	S'2	T' T'
25/07/66	A	X			X						X				X								X		X	X			X
25/10/66	A	X			X						X				X								X		X	X			X
26/01/67	B	X			X				X		X	X			X								X		X	X			X
26/04/67	B	X			X						X				X								X		X	X			X
26/07/67	B	X			X						X				X								X		X	X			X
26/10/67	B	X			X						X				X								X		X	X			X
25/01/68	B	X			X						X				X								X		X	X			X
30/04/68	B	X			X						X	X			X				X				X		X	X			X
30/07/68	B	X			X						X	X			X				X				X		X	X			X
31/10/68	B	X			X						X	X			X				X				X		X				X
30/01/69	B	X			X		X				X	X			X				X				X		X				X
27/03/69	B	X			X		X				X	X			X				X				X		X				X
25/06/69	B	X			X		X				X	X			X				X				X		X				X
04/08/69	B	X			X		X				X	X			X				X				X		X				X
05/11/69	C	X			X		X				X	X			X				X				X		X				X
05/02/70	C	X			X		X				X	X			X				X				X		X				X
05/05/70	C	X			X						X	X			X				X				X		X				X
05/08/70	C	X			X		X				X	X			X				X				X		X				
04/11/70	C	X			X		X				X	X			X				X				X		X				
04/03/71	C	X			X		X				X	X			X				X				X		X				
10/06/71	C	X			X		X				X	X			X				X				X		X	X			
27/07/71	D	X			X		X				X	X			X				X				X		X	X			
28/10/71	D	X			X		X				X	X			X		X		X				X		X	X			
28/02/72	D	X			X		X				X	X			X		X		X				X		X	X			
29/06/72	D	X			X		X				X	X			X		X		X				X		X	X			
26/10/72	D	X			X		X				X	X			X		X		X				X		X	X			
28/02/73	D	X			X		X				X	X			X		X		X				X		X	X			
18/04/73	E	X			X		X				X	X			X		X		X				X		X	X			
21/08/73	E	X			X		X				X	X			X		X		X				X		X	X			
20/12/73	E	X			X		X				X	X			X		X		X				X		X	X			
23/04/74	E	X			X		X					X			X		X		X						X	X			
13/05/74	E	X			X		X					X			X		X		X						X	X			
14/08/74	F	X			X		X					X			X		X		X						X	X			
12/12/74	F	X			X		X				X	X			X		X		X						X	X			
15/04/75	F	X			X		X				X	X			X		X		X						X	X			
15/05/75	G	X			X		X				X	X			X		X		X						X	X			
19/08/75	G	X			X		X				X	X			X		X		X						X	X			
17/12/75	G	X			X		X				X	X			X		X		X				X		X	X			
18/03/76	G	X			X		X					X			X		X		X				X		X	X			
04/05/76	G	X			X		X				X	X			X		X		X				X		X	X			
05/08/76	G	X			X		X				X	X			X		X		X				X		X	X			
04/11/76	H	X			X		X				X	X			X		X		X				X		X	X			
05/03/77	H	X			X		X				X	X			X		X		X				X		X	X			

Date	System Lab Nos	F F1	F2	V1	V2	N'1	N'2	V'	J J1	J2	L L	M M1	M2	M'	S S1	S2	U1	U2	H'	U'1	U'2	H"	S"	U"	Z Z	R' R'	S'1	S'2	T' T'
20/04/77	H	X			X		X				X				X		X		X				X		X	X			
20/07/77	K	X			X		X				X				X		X		X				X		X	X			
20/10/77	K	X			X		X				X				X		X		X				X		X	X			
24/01/78	K	X			X		X								X		X		X				X		X	X			
20/04/78	K	X			X		X				X				X		X		X				X		X	X			
10/05/78	K	X			X		X				X				X		X		X				X		X	X			
10/08/78	L	X			X		X				X				X		X		X				X		X	X			
08/11/78	L	X			X		X				X				X		X		X				X		X	X			
12/02/79	L	X			X		X				X				X		X		X				X		X	X			
10/05/79	L	X			X		X								X		X		X				X		X	X			
12/06/79	L	X			X		X								X		X		X				X		X	X			
15/10/79	L		X		X		X				X				X		X		X				X		X	X			
14/02/80	L	X			X		X				X				X		X						X			X			
24/04/80	M	X			X		X				X				X		X						X			X			
27/08/80	M	X			X		X				X				X		X						X			X			
23/12/80	M	X			X		X								X		X						X		X				
09/02/81	M	X			X		X				X				X		X						X		X				
09/06/81	M	X			X		X								X		X						X		X				
07/10/81	M	X			X		X				X	X			X		X					X			X				
10/12/81	N	X			X		X		X		X	X			X		X						X		X				
06/04/82	N	X			X		X				X				X		X								X				
05/08/82	N	X			X		X				X				X		X								X				
15/09/82	N	X		X		X			X		X	X			X		X		X						X				
15/12/82	N	X		X		X					X	X			X		X		X			X	X		X	X			
16/03/83	N	X		X		X					X	X			X		X		X			X	X		X	X			
06/04/83	P	X		X		X	X				X				X		X		X	X		X	X	X					
06/07/83	P	X		X		X			X		X	X			X		X		X				X		X	X			
06/10/83	P	X			X	X			X		X	X			X		X		X				X		X	X			
16/11/83	P	X		X		X			X		X	X			X		X		X				X		X	X			
11/04/84	P	X		X		X	X				X				X		X		X	X		X	X	X					
13/07/84	P	X		X		X	X				X	X			X		X		X	X		X	X	X	X	X			
12/10/84	P	X		X		X	X				X	X			X		X		X	X		X	X	X	X	X			X
15/01/85	P	X		X		X	X				X	X			X		X		X			X		X	X	X			
13/02/85	Q	X			X	X	X				X	X			X		X		X	X		X		X	X	X			
16/05/85	Q	X			X	X	X				X	X			X		X		X	X		X			X	X			
15/08/85	Q	X			X	X	X				X	X			X		X		X	X		X			X	X			
15/11/85	Q	X			X	X	X				X	X			X		X		X	X		X			X	X			
14/02/86	Q	X			X	X	X				X	X			X		X		X	X		X			X	X			
15/05/86	Q	X			X	X	X				X				X		X		X	X					X	X			
15/08/86	R	X			X	X	X				X	X					X		X	X					X	X			
19/11/86	R	X			X	X	X				X	X					X		X	X					X	X			
17/02/87	R	X			X		X				X	X					X		X	X					X	X			
15/05/87	R	X			X		X				X	X			X		X	X	X						X	X			

Date	System Lab Nos	F F1	F2	V1	V2	N'1	N'2	V'	J J1	J2	L L	M M1	M2	M'	S S1	S2	U1	U2	H'	U'1	U'2	H"	S"	U"	Z Z	R' R'	S'1	S'2	T' T'
01/07/87	R	X			X		X				X	X			X		X	X	X				X		X	X			
21/08/87	R	X			X		X				X	X			X		X	X	X				X		X	X			
25/11/87	S	X			X		X				X	X			X		X	X	X				X		X	X			
29/01/88	S	X			X		X				X	X			X		X	X	X				X		X	X			
25/02/88	S	X			X		X				X	X			X		X	X	X			X	X		X				
27/05/88	S	X			X		X				X	X			X		X	X	X				X		X				
18/08/88	S	X			X		X				X	X			X		X	X	X			X	X	X	X				
24/11/88	S	X			X		X				X	X			X		X	X	X			X			X				
17/02/89	T	X			X		X				X						X	X	X						X				
25/05/89	T	X			X		X				X						X	X	X						X				
25/08/89	T	X			X		X				X						X	X	X						X	X			
24/11/89	T	X			X		X				X						X	X	X						X	X			
23/02/90	V	X	X		X		X				X						X	X	X						X				
25/05/90	V	X	X		X		X				X						X	X	X						X				
23/08/90	V	X	X		X		X				X						X	X	X						X				
24/11/90	V	X	X	X	X		X				X						X	X	X						X				
22/02/91	V	X	X		X		X				X	X					X	X	X						X				
24/05/91	V	X	X		X		X				X	X					X	X	X						X				
23/08/91	V	X	X		X		X				X	X					X	X	X						X				
22/11/91	V	X	X		X		X				X	X					X	X	X						X				
21/02/92	V	X	X		X		X				X	X			X		X		X						X				
22/05/92	V		X		X		X				X	X			X		X		X						X				
21/08/92	V	X			X		X				X	X			X		X	X	X		X				X				
20/11/92	V19		X	X			X				X	X			X		X		X						X				
24/02/93	V20		X	X			X				X	X			X		X		X						X				
26/05/93	V21		X	X			X				X	X			X		X		X						X				
25/08/93	V23		X	X			X				X	X			X		X		X						X				
26/11/93	V25	X	X	X			X				X	X			X		X		X						X				
25/02/94	V27	X	X	X			X				X	X			X		X		X						X				
25/05/94	V28	X		X			X				X				X		X		X						X				
24/08/94	V30	X		X			X				X						X		X						X				
23/11/94	V32	X		X			X				X				X		X		X						X				
22/02/95	V34	X		X			X				X						X		X						X				
24/05/95	V36	X		X			X				X						X		X						X				
25/08/95	V38	X	X		X		X				X							X	X						X				
24/11/95	V40	X	X	X			X		X		X							X	X						X				
21/02/96	V42	X	X	X			X		X		X							X	X						X				
29/05/96	V44	X			X		X				X	X			X			X	X						X				
28/08/96	V46				X		X				X				X	X	X		X						X				
20/11/96	V49	X		X	X		X				X				X	X	X		X						X				
05/03/97	V51		X	X	X		X				X				X	X	X		X						X				

Appendix II

Allele frequencies at the blood type systems A, B, C, F, L, S, Z and the serum protein loci transferrin and albumin

Frequencies in thirty-seven European cattle breeds and eighteen Hereford cattle populations

Blood type system	A	C				L	S			Z
Antigenic factor	A	C	R	W	X	L	S	U	H'	Z
Aberdeen Angus	0.387	0.603	0.013	0.400	0.099	0.074	0.107	0.008	0.395	0.288
Ayrshire	0.454	0.497		0.419	0.270	0.194	0.030		0.392	0.253
Belgian Blue	0.307	0.515	0.037	0.359	0.168	0.137	0.078	0.013	0.438	0.143
Belted Galloway	0.233	0.669	0.007	0.328	0.101	0.101	0.140		0.427	0.071
Blonde d'Aquitaine	0.737	0.329	0.252	0.656	0.286	0.106	0.203	0.277	0.500	0.490
British White	0.469	0.709	0.036	0.383	0.043	0.249	0.073	0.144	0.259	0.222
Brown Swiss	0.283	0.679	0.109	0.500	0.134		0.134	0.007	0.486	0.196
Charolais	0.721	0.468	0.039	0.590	0.089	0.195	0.127	0.404	0.638	0.502
Chianina	0.574	0.523	0.222	0.787	0.385	0.087	0.174	0.282	0.826	0.592
Dexter	0.289	0.546	0.018	0.330	0.121	0.245	0.016	0.006	0.442	0.420
Galloway	0.240	0.414	0.032	0.209	0.090	0.293	0.073	0.008	0.339	0.116
Gelbvieh	0.529	0.667	0.275	0.826	0.357	0.209	0.057	0.005	0.826	0.539
Gloucester	0.432	0.748	0.032	0.096	0.111	0.602	0.032	0.003	0.076	0.166
Guernsey	0.570	0.493	0.063	0.435	0.212	0.374	0.268		0.502	0.443
Hereford (horned)	0.517	0.521	0.466	0.377	0.023	0.532	0.258	0.002	0.274	0.367
Hereford (poll)	0.507	0.578	0.308	0.291	0.055	0.469	0.148	0.003	0.238	0.339
Highland	0.275	0.813		0.216	0.131	0.063	0.054		0.470	0.045
Icelandic	0.138	0.969		0.385	0.166	0.385			0.889	0.587
Holstein Friesian	0.361	0.255	0.025	0.173	0.222	0.178	0.086	0.033	0.403	0.189
Jersey	0.918	0.225	0.096	0.819	0.260	0.247	0.230	0.016	0.694	0.347
Kerry	0.861	0.520	0.029	0.180	0.253	0.203			0.335	0.321
Limousin	0.680	0.358	0.049	0.682	0.247	0.118	0.090	0.255	0.488	0.189
Lincoln Red	0.472	0.969		0.323	0.051	0.059	0.008		0.538	0.017
Longhorn	0.969	0.826	0.031	0.251	0.055	0.410	0.326	0.104	0.349	0.478
Maine-Anjou	0.518	0.569	0.012	0.215	0.208	0.060	0.012	0.098	0.440	0.124
Marchigiana	0.454	0.803	0.020	0.386	0.376	0.040	0.202	0.430	0.839	0.622
Meuse Rhine Yssel	0.405	0.468	0.079	0.297	0.262	0.165	0.084		0.450	0.255
Murray Grey	0.323	0.714		0.095	0.151	0.113	0.059	0.042	0.427	0.051
Piemontese	0.398	0.410	0.091	0.504	0.184	0.149	0.060	0.091	0.435	0.550
Romagnola	0.680	0.266	0.390	0.774	0.723	0.101	0.380	0.130	0.700	0.507
Salers	0.783	0.167	0.042	0.675	0.086	0.126	0.203	0.358	0.758	0.491
Shorthorn	0.549	0.694	0.032	0.339	0.143	0.261	0.090		0.282	0.293
Simmental	0.366	0.334	0.039	0.633	0.070	0.120	0.175	0.023	0.669	0.460
South Devon	0.512	0.969	0.012	0.711	0.081	0.114	0.127	0.006	0.260	0.081
Sussex	0.669	0.200	0.040	0.099	0.171	0.646	0.040	0.125	0.315	0.293
Welsh Black	0.405	0.672		0.418	0.288	0.504	0.039	0.008	0.432	0.380
White Park	0.140	0.745	0.045	0.293	0.091	0.011	0.045	0.022	0.324	0.033

Blood type system	B																	
Antigenic factor	B	G	I	K	O	P	Q	T	Y	A'	D'	E'	G'	I'	K'	O'	P'	Q'
Aberdeen Angus	0.110	0.142	0.119		0.317	0.026	0.034	0.034	0.383	0.079	0.060	0.439	0.082	0.178	0.005	0.052	0.082	0.160
Ayrshire	0.170	0.206	0.007		0.503	0.003	0.043	0.023	0.321	0.163	0.114	0.020	0.088	0.262	0.020	0.114	0.007	0.085
Belgian Blue	0.327	0.399	0.091	0.015	0.314	0.007	0.007	0.001	0.378	0.155	0.105	0.215	0.138	0.179	0.006	0.159	0.081	0.124
Belted Galloway	0.181	0.140	0.021	0.007	0.215		0.014	0.101	0.172	0.101	0.014	0.233	0.042	0.148		0.164	0.172	0.116
Blonde d'Aquitaine	0.288	0.185	0.214	0.003	0.351	0.010	0.197	0.157	0.331	0.349	0.380	0.113	0.149	0.220	0.191	0.248	0.173	0.051
British White	0.128	0.144		0.014	0.178	0.007	0.007	0.021	0.043	0.029		0.021	0.029	0.169	0.007	0.014	0.029	0.288
Brown Swiss	0.444	0.283	0.093	0.061	0.471	0.101	0.045	0.303	0.790	0.178	0.084	0.515	0.444	0.093	0.015	0.143	0.272	0.272
Charolais	0.296	0.133	0.014	0.010	0.238	0.227	0.048	0.090	0.174	0.268	0.087	0.061	0.213	0.402	0.011	0.160	0.209	0.058
Chianina	0.272	0.104	0.096		0.969		0.202	0.464	0.231	0.130		0.015	0.112	0.071	0.293	0.165	0.304	0.047
Dexter	0.242	0.221	0.021		0.234	0.014	0.080	0.198	0.330	0.125	0.125	0.208	0.289	0.377	0.014	0.100	0.221	0.348
Galloway	0.190	0.171	0.073	0.056	0.090	0.008	0.032	0.040	0.304	0.040	0.032	0.363	0.056	0.315	0.065	0.099	0.008	0.134
Gelbvieh	0.118	0.084	0.184	0.107	0.241	0.005	0.130	0.010	0.124	0.073	0.052	0.202	0.297	0.090	0.005	0.202		0.153
Gloucester	0.225	0.132	0.349	0.111	0.529		0.006	0.042	0.449	0.212		0.045	0.344	0.136	0.455	0.449	0.062	0.083
Guernsey	0.404	0.244	0.363	0.075	0.349	0.018	0.034	0.115	0.189	0.154	0.027	0.316	0.151	0.415	0.271	0.271	0.075	0.164
Hereford (horned)	0.009	0.011	0.001	0.006	0.217	0.013	0.092	0.001	0.555	0.250	0.364	0.003	0.004	0.522		0.010	0.011	0.115
Hereford (poll)	0.011	0.009	0.004	0.004	0.214	0.011	0.050		0.511	0.181	0.387	0.012	0.007	0.570		0.023	0.010	0.096
Highland	0.009	0.054			0.092		0.027	0.121	0.470	0.325	0.152	0.036	0.152	0.275		0.082	0.009	0.036
Holstein Friesian	0.133	0.284	0.125	0.004	0.230	0.023	0.003	0.003	0.391	0.123	0.097	0.271	0.138	0.103	0.084	0.153	0.040	0.153
Icelandic	0.159	0.096	0.276	0.037	0.227	0.050			0.025	0.189		0.366	0.152	0.219				0.227
Jersey	0.529	0.518	0.004	0.129	0.369	0.004	0.033	0.272	0.462	0.381	0.160	0.034	0.442	0.270	0.575	0.498	0.058	0.100
Kerry	0.253	0.203	0.191		0.660	0.019			0.584	0.365	0.168	0.216	0.080	0.091		0.029	0.112	0.365
Limousin	0.561	0.234	0.168	0.009	0.305	0.016	0.203	0.204	0.132	0.283	0.180	0.298	0.149	0.387	0.091	0.290	0.180	0.162
Longhorn	0.055	0.315	0.008	0.015	0.138	0.087	0.023	0.008	0.877	0.008	0.047	0.337	0.015	0.315	0.039	0.031	0.079	0.202
Lincoln Red	0.211	0.335			0.778		0.008		0.521	0.472	0.042	0.373	0.170	0.059		0.347	0.557	0.347
Maine-Anjou	0.269	0.269	0.066	0.012	0.642	0.137	0.054	0.079	0.293	0.277	0.006	0.079	0.124	0.208	0.079	0.144	0.137	0.193
Marchigiana	0.262	0.088	0.047	0.007	0.803	0.074	0.186	0.503	0.186	0.442	0.047	0.088	0.060	0.170	0.244	0.096	0.408	0.194
Meuse Rhine Yssel	0.202	0.135	0.057		0.397		0.005		0.255	0.304	0.159	0.118	0.047	0.222	0.052	0.020	0.036	0.068
Murray Grey	0.113	0.017	0.200		0.211		0.017	0.051	0.311	0.141	0.019	0.232	0.190	0.335		0.077	0.095	0.311
Piemontese	0.149	0.211	0.099	0.015	0.220	0.099	0.115	0.220	0.239	0.132	0.157	0.288	0.175	0.149	0.258	0.220	0.075	0.124
Romagnola	0.533	0.321	0.087	0.080	0.533	0.073	0.370	0.370	0.160	0.153	0.138	0.494	0.249	0.145	0.184	0.232	0.370	0.138
Salers	0.233	0.233	0.386		0.181	0.036	0.112	0.119	0.093	0.249	0.139	0.225	0.061	0.503	0.106	0.132	0.018	0.086
Shorthorn	0.401	0.065	0.032	0.008	0.375	0.016			0.455	0.271	0.152	0.143	0.327	0.209	0.032	0.048	0.339	0.200
Simmental	0.181	0.127	0.108	0.011	0.220	0.020	0.087	0.051	0.094	0.179	0.037	0.150	0.142	0.298	0.050	0.148	0.028	0.100
South Devon	0.213	0.345	0.074	0.006	0.244		0.120	0.114	0.293	0.107	0.107	0.293	0.094	0.134	0.074	0.087	0.134	0.107
Sussex	0.116	0.016			0.240	0.016	0.056	0.040	0.401	0.065	0.056	0.375	0.065	0.455	0.073	0.016	0.032	0.427
Welsh Black	0.123	0.235	0.123	0.008	0.380	0.055	0.008	0.023	0.332	0.187	0.016	0.080	0.114	0.177	0.023	0.177	0.159	0.159
White Park	0.033	0.357		0.011	0.234	0.011		0.033	0.745	0.166		0.341	0.033	0.308		0.056	0.068	0.166

Locus	F		Transferrin				Albumin	
Allele	F	V	A	D1	D2	E	A	B
Aberdeen Angus	0.711	0.289	0.500	0.035	0.329	0.135	0.994	0.006
Ayrshire	0.860	0.140	0.370	0.156	0.339	0.135	0.990	0.010
Belgian Blue	0.936	0.064	0.412	0.216	0.242	0.129	0.860	0.140
Belted Galloway	0.795	0.206	0.421	0.211	0.290	0.079	1.000	
Blonde d'Aquitaine	0.648	0.352	0.397	0.180	0.423		0.988	0.012
British White	0.972	0.028	0.359	0.039	0.603		1.000	
Brown Swiss	0.735	0.265	0.118	0.088	0.618	0.177	0.962	0.039
Charolais	0.739	0.261	0.294	0.146	0.553	0.007	0.849	0.151
Chianina	0.758	0.242	0.583	0.333	0.083		0.917	0.083
Dexter	0.800	0.200	0.302	0.194	0.260	0.245	0.998	0.002
Galloway	0.891	0.109	0.231	0.154	0.192	0.423	1.000	
Gelbvieh	0.732	0.268	0.306	0.218	0.453	0.024	0.977	0.023
Gloucester	0.642	0.358	0.713	0.105	0.183		0.977	0.023
Guernsey	0.815	0.185	0.509	0.312	0.179		0.991	0.009
Hereford (horned)	0.851	0.149	0.397	0.233	0.368	0.002	1.000	
Hereford (poll)	0.806	0.194	0.513	0.147	0.334	0.005	0.997	0.003
Highland	0.851	0.149	0.565	0.129	0.194	0.113	0.936	0.065
Holstein Friesian	0.825	0.176	0.380	0.178	0.412	0.030	0.992	0.008
Icelandic	1.000		0.293	0.518	0.177	0.012	1.000	
Jersey	0.739	0.261	0.748	0.058	0.195		0.998	0.003
Kerry	0.615	0.385	0.510	0.059	0.304	0.128	1.000	
Limousin	0.779	0.221	0.221	0.310	0.461	0.008	1.000	
Longhorn	0.599	0.402	0.115	0.077	0.808		1.000	
Lincoln Red	0.951	0.049	0.435	0.065	0.283	0.217	1.000	
Maine-Anjou	0.605	0.395	0.250	0.200	0.500	0.050	0.938	0.063
Marchigiana	0.766	0.234	0.500	0.083	0.375	0.042	1.000	
Meuse Rhine Yssel	0.859	0.141	0.583	0.194	0.208	0.014	1.000	
Murray Grey	0.697	0.303	0.313	0.250	0.375	0.063	1.000	
Piemontese	0.797	0.203	0.452	0.191	0.357		1.000	
Romagnola	0.628	0.372	0.516	0.141	0.328	0.016	0.774	0.226
Salers	0.671	0.329	0.179	0.321	0.500		0.929	0.071
Shorthorn	0.711	0.289	0.457	0.304	0.239		1.000	
Simmental	0.793	0.207	0.202	0.142	0.569	0.087	0.978	0.022
South Devon	0.923	0.077	0.172	0.414	0.414		0.950	0.050
Sussex	0.969	0.031	0.473	0.419	0.108		0.982	0.019
Welsh Black	0.739	0.262	0.415	0.053	0.511	0.021	1.000	
White Park	0.891	0.109	0.273		0.205	0.523	1.000	

Bloodtype system	A	C				L	S			Z
Antigenic factor	**A**	**C**	**R**	**W**	**X**	**L**	**S**	**U**	**H'**	**Z**
Hereford populations										
Traditional	0.457	0.557	0.488	0.457		0.521	0.095		0.151	0.400
British 1960-69	0.468	0.524	0.648	0.406	0.024	0.627	0.271		0.282	0.406
British 1970-79	0.621	0.530	0.414	0.399	0.024	0.393	0.292		0.307	0.375
British 1980-89	0.520	0.804	0.481	0.481		0.660	0.216		0.266	0.293
"Hybrid" 1970-79	0.558	0.558	0.308	0.308	0.022	0.410	0.192		0.206	0.324
"Hybrid" 1980-89	0.435	0.491	0.289	0.276	0.021	0.628	0.179	0.006	0.260	0.268
"Hybrid" 1990-96	0.287	0.771	0.522	0.216	0.027	0.735	0.036		0.251	0.251
British Poll1960-69	0.374	0.527	0.463	0.246	0.062	0.654	0.175		0.296	0.374
British Poll 1970-79	0.597	0.555	0.300	0.304	0.063	0.351	0.171	0.001	0.245	0.331
British Poll 1980-89	0.480	0.649	0.292	0.307	0.042	0.562	0.101	0.004	0.240	0.338
British Poll 1990-96	0.296	0.708	0.453	0.333	0.009	0.642	0.039		0.243	0.408
Canadian 1960-79	0.368	0.800	0.225	0.152	0.020	0.471	0.062		0.279	0.175
Canadian 1980-90	0.309	0.723	0.321	0.288	0.031	0.489	0.016		0.277	0.288
Canadian Poll 1960-79	0.481	0.969	0.412	0.445	0.019	0.445	0.039		0.240	0.293
Canadian Poll 1980-90	0.520	0.969	0.445	0.445	0.053	0.380			0.184	0.321
Irish	0.281	0.588	0.319	0.314	0.029	0.448	0.071	0.007	0.295	0.408
New Zealand	0.374	0.349	0.237	0.229	0.185	0.330	0.086	0.003	0.217	0.316
Swedish	0.375	0.620	0.467	0.257	0.078	0.671	0.144	0.339	0.242	0.237

Bloodtype System	B																	
Antigenic factor	**B**	**G**	**I**	**K**	**O**	**P**	**Q**	**T**	**Y**	**A'**	**D'**	**E'**	**G'**	**I'**	**K'**	**O'**	**P'**	**Q'**
Hereford populations																		
Traditional	0.008				0.160		0.017		0.686	0.323	0.360			0.538				0.042
British 1960-69	0.010	0.010		0.009	0.154	0.017	0.136		0.540	0.254	0.362	0.001	0.003	0.538		0.010	0.005	0.115
British 1970-79	0.004	0.002			0.315	0.011	0.075		0.554	0.256	0.369		0.006	0.514		0.002	0.017	0.122
British 1980-89		0.039			0.293		0.019		0.660	0.321	0.481	0.039		0.562				0.060
"Hybrid" 1970-79	0.022	0.022		0.022	0.293	0.022	0.079		0.670	0.308	0.468	0.011	0.011	0.489		0.022		0.166
"Hybrid" 1980-89	0.037	0.040	0.003	0.006	0.236		0.049		0.534	0.224	0.289	0.003	0.003	0.457		0.031	0.012	0.088
"Hybrid" 1990-96	0.194	0.263		0.228	0.287				0.312	0.228	0.194			0.487		0.131		0.009
British Poll1960-69	0.028	0.028	0.004	0.028	0.185	0.020	0.045		0.571	0.185	0.361	0.033		0.571		0.028	0.008	0.088
British Poll 1970-79	0.008	0.006	0.004	0.001	0.264	0.011	0.052		0.475	0.209	0.339	0.012	0.002	0.580		0.015	0.017	0.086
British Poll 1980-89	0.007	0.005	0.004		0.171	0.004	0.048		0.509	0.140	0.439	0.013	0.017	0.578		0.035	0.004	0.107
British Poll 1990-96	0.044	0.066		0.026	0.210	0.017	0.026		0.511	0.216	0.401	0.022		0.587		0.035	0.009	0.048
Canadian 1960-79	0.062	0.062		0.041	0.200	0.020	0.062		0.510	0.106	0.400		0.020	0.600		0.062	0.020	0.062
Canadian 1980-90	0.080	0.089		0.023	0.114	0.016	0.023		0.474	0.187	0.368	0.008	0.039	0.405		0.080	0.008	0.023
Canadian Poll 1960-79					0.168		0.039		0.520	0.101	0.520			0.608			0.039	0.019
Canadian Poll 1980-90	0.053	0.066		0.013	0.168		0.066		0.642	0.123	0.469		0.013	0.520		0.080		0.039
Irish	0.043	0.059	0.003	0.007	0.267	0.009	0.060	0.002	0.387	0.252	0.285	0.012	0.021	0.398	0.001	0.035	0.005	0.123
New Zealand	0.039	0.029	0.003	0.013	0.288	0.022	0.032	0.003	0.466	0.204	0.334	0.010	0.006	0.503		0.032	0.013	0.052
Swedish	0.108	0.104	0.001	0.074	0.016	0.004	0.085		0.507	0.239	0.306	0.011	NA	0.396	0.002	0.121	0.002	0.108

Appendix III

Squared gene frequency correlations (linkage disequilibria) between pairs of B system factors in nine cattle breeds

Matrices of correlations are symmetric, and values greater than 0.1 are shaded

Blonde d'Aquitaine

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.000	0.004	0.013	0.000	0.074	0.013	0.011	0.000	0.000	0.029	0.019	0.007	0.005	0.003	0.001	0.004	0.000	0.000	0.007	0.002
G_1	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.006	0.003	0.000	0.036	0.000	0.000	0.000	0.069	0.000	0.000	0.000
I_1	0.004	0.000	0.000	0.000	0.012	0.006	0.000	0.003	0.000	0.015	0.000	0.000	0.003	0.003	0.066	0.002	0.050	0.000	0.004	0.004	0.002
I_2	0.013	0.000	0.000	0.000	0.000	0.006	0.000	0.011	0.000	0.001	0.006	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K_1	0.000	0.000	0.012	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.013	0.000	0.000	0.000	0.001	0.000	0.000	0.062	0.009
O_1	0.074	0.001	0.006	0.006	0.001	0.000	0.006	0.000	0.000	0.023	0.062	0.042	0.016	0.014	0.033	0.001	0.233	0.000	0.004	0.003	0.002
P_1	0.013	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.001	0.006	0.001	0.000	0.036	0.045	0.000	0.000	0.000	0.033	0.000	0.000
Q_2	0.011	0.000	0.003	0.011	0.001	0.000	0.000	0.000	0.000	0.011	0.030	0.093	0.000	0.045	0.006	0.002	0.004	0.017	0.037	0.000	0.004
T_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Y_2	0.000	0.001	0.015	0.001	0.001	0.023	0.001	0.011	0.000	0.000	0.001	0.072	0.019	0.001	0.006	0.000	0.099	0.009	0.001	0.009	0.066
A'_1	0.029	0.006	0.000	0.006	0.000	0.062	0.006	0.030	0.000	0.001	0.000	0.013	0.027	0.002	0.000	0.002	0.001	0.008	0.006	0.023	0.001
D'	0.019	0.003	0.000	0.003	0.001	0.042	0.001	0.093	0.000	0.072	0.013	0.000	0.002	0.003	0.022	0.000	0.020	0.003	0.004	0.021	0.000
F'_1	0.007	0.000	0.003	0.000	0.013	0.016	0.000	0.000	0.000	0.019	0.027	0.002	0.000	0.071	0.002	0.001	0.012	0.021	0.029	0.025	0.003
G'_1	0.005	0.036	0.003	0.000	0.000	0.014	0.036	0.045	0.000	0.001	0.002	0.003	0.071	0.000	0.011	0.010	0.011	0.149	0.313	0.010	0.004
I'_1	0.003	0.000	0.066	0.000	0.000	0.033	0.045	0.006	0.000	0.006	0.000	0.022	0.002	0.011	0.000	0.014	0.028	0.004	0.000	0.030	0.008
J'_1	0.001	0.000	0.002	0.000	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.000	0.001	0.010	0.014	0.000	0.037	0.208	0.008	0.007	0.000
K'	0.004	0.000	0.050	0.000	0.001	0.233	0.000	0.004	0.000	0.099	0.001	0.020	0.012	0.011	0.028	0.037	0.000	0.002	0.005	0.000	0.014
O'_1	0.000	0.069	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.009	0.008	0.003	0.021	0.149	0.004	0.208	0.002	0.000	0.022	0.020	0.000
P'_1	0.000	0.000	0.004	0.000	0.000	0.004	0.033	0.037	0.000	0.001	0.006	0.004	0.029	0.313	0.000	0.008	0.005	0.022	0.000	0.008	0.007
Q'	0.007	0.000	0.004	0.000	0.062	0.003	0.000	0.000	0.000	0.009	0.023	0.021	0.025	0.010	0.030	0.007	0.000	0.020	0.008	0.000	0.002
Y'	0.002	0.000	0.002	0.000	0.009	0.002	0.000	0.004	0.000	0.066	0.001	0.000	0.003	0.004	0.008	0.000	0.014	0.000	0.007	0.002	0.000

Belgian Blue

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.025	0.027	0.000	0.008	0.025	0.001	0.000	0.000	0.001	0.000	0.000	0.003	0.009	0.000	0.000	0.005	0.007	0.005	0.018	0.009
G_1	0.025	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.025	0.003	0.000	0.004	0.001	0.000	0.000	0.000	0.002	0.003	0.001	0.000
I_1	0.027	0.001	0.000	0.001	0.000	0.014	0.000	0.013	0.000	0.004	0.007	0.005	0.003	0.007	0.000	0.000	0.009	0.003	0.001	0.004	0.003
I_2	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.004	0.000	0.000	0.000	0.000	0.076	0.018	0.000	0.000	0.000	0.000
K_1	0.008	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.005	0.016	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
O_1	0.025	0.002	0.014	0.001	0.001	0.000	0.000	0.003	0.000	0.001	0.000	0.103	0.006	0.019	0.001	0.000	0.005	0.007	0.002	0.022	0.003
P_1	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.001	0.001	0.000	0.000	0.000	0.002	0.000	0.051	0.000	0.000	0.000	0.050
Q_2	0.000	0.000	0.013	0.000	0.000	0.003	0.005	0.000	0.000	0.002	0.007	0.000	0.002	0.001	0.008	0.000	0.015	0.000	0.000	0.001	0.008
T_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Y_2	0.001	0.025	0.004	0.001	0.005	0.001	0.001	0.002	0.000	0.000	0.034	0.054	0.001	0.005	0.007	0.001	0.002	0.010	0.010	0.014	0.001
A'_1	0.000	0.003	0.007	0.004	0.016	0.000	0.001	0.007	0.000	0.034	0.000	0.012	0.003	0.054	0.004	0.003	0.002	0.166	0.094	0.001	0.014
D'	0.000	0.000	0.005	0.000	0.001	0.103	0.000	0.000	0.000	0.054	0.012	0.000	0.132	0.007	0.003	0.000	0.006	0.003	0.001	0.014	0.000
F'_1	0.003	0.004	0.003	0.000	0.000	0.006	0.000	0.002	0.000	0.001	0.003	0.132	0.000	0.000	0.001	0.000	0.000	0.001	0.004	0.000	0.000
G'_1	0.009	0.001	0.007	0.000	0.000	0.019	0.000	0.001	0.000	0.005	0.054	0.007	0.000	0.000	0.003	0.000	0.000	0.069	0.056	0.019	0.000
I'_1	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.008	0.000	0.007	0.004	0.003	0.001	0.003	0.000	0.000	0.001	0.001	0.011	0.036	0.012
J'_1	0.000	0.000	0.000	0.076	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.011	0.000	0.000
K'	0.005	0.000	0.009	0.018	0.000	0.005	0.051	0.015	0.000	0.002	0.002	0.006	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.036
O'_1	0.007	0.002	0.003	0.000	0.000	0.007	0.000	0.000	0.000	0.010	0.166	0.003	0.001	0.069	0.001	0.008	0.000	0.000	0.069	0.002	0.000
P'_1	0.005	0.003	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.010	0.094	0.001	0.004	0.056	0.011	0.011	0.000	0.069	0.000	0.027	0.000
Q'	0.018	0.001	0.004	0.000	0.001	0.022	0.000	0.001	0.000	0.014	0.001	0.014	0.000	0.019	0.036	0.000	0.001	0.002	0.027	0.000	0.000
Y'	0.009	0.000	0.003	0.000	0.000	0.003	0.050	0.008	0.000	0.001	0.014	0.000	0.000	0.000	0.012	0.000	0.036	0.000	0.000	0.000	0.000

Charolais

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.012	0.002	0.000	0.001	0.065	0.002	0.000	0.008	0.006	0.000	0.024	0.001	0.009	0.002	0.000	0.006	0.004	0.006	0.004	0.000
G_1	0.012	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.007	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.001
I_1	0.002	0.015	0.000	0.000	0.010	0.001	0.005	0.008	0.000	0.001	0.000	0.004	0.000	0.000	0.004	0.003	0.151	0.004	0.000	0.005	0.053
I_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.238	0.000	0.006	0.000	0.000	0.000
K_1	0.001	0.000	0.010	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.001	0.007
O_1	0.065	0.000	0.001	0.000	0.000	0.000	0.000	0.002	0.000	0.062	0.011	0.073	0.025	0.001	0.005	0.000	0.002	0.003	0.000	0.003	0.000
P_1	0.002	0.000	0.005	0.000	0.000	0.000	0.000	0.001	0.000	0.014	0.002	0.005	0.000	0.000	0.296	0.004	0.010	0.000	0.002	0.000	0.001
Q_2	0.000	0.000	0.008	0.001	0.002	0.002	0.001	0.000	0.000	0.005	0.000	0.016	0.042	0.000	0.005	0.000	0.042	0.025	0.042	0.000	0.008
T_1	0.008	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.010	0.002	0.005	0.000	0.000	0.000	0.009	0.000	0.000
Y_2	0.006	0.001	0.001	0.000	0.000	0.062	0.014	0.005	0.000	0.000	0.016	0.099	0.003	0.003	0.004	0.000	0.001	0.008	0.000	0.001	0.025
A'_1	0.000	0.001	0.000	0.000	0.000	0.011	0.002	0.000	0.001	0.016	0.000	0.000	0.002	0.041	0.035	0.000	0.004	0.014	0.054	0.011	0.003
D'	0.024	0.001	0.004	0.000	0.010	0.073	0.005	0.016	0.000	0.099	0.000	0.000	0.001	0.010	0.014	0.000	0.019	0.001	0.012	0.003	0.022
F'_1	0.001	0.000	0.000	0.001	0.000	0.025	0.000	0.042	0.010	0.003	0.002	0.001	0.000	0.017	0.001	0.000	0.003	0.000	0.005	0.003	0.000
G'_1	0.009	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.002	0.003	0.041	0.010	0.017	0.000	0.000	0.000	0.002	0.008	0.101	0.143	0.008
I'_1	0.002	0.000	0.004	0.000	0.000	0.005	0.296	0.005	0.005	0.004	0.035	0.014	0.001	0.000	0.000	0.002	0.001	0.009	0.006	0.006	0.000
J'_1	0.000	0.000	0.003	0.238	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.003	0.008	0.000	0.000	0.002
K'	0.006	0.000	0.151	0.000	0.038	0.002	0.010	0.042	0.000	0.001	0.004	0.019	0.003	0.002	0.001	0.003	0.000	0.005	0.001	0.017	0.074
O'_1	0.004	0.002	0.004	0.006	0.000	0.003	0.000	0.025	0.000	0.008	0.014	0.001	0.000	0.008	0.009	0.008	0.005	0.000	0.045	0.002	0.006
P'_1	0.006	0.001	0.000	0.000	0.000	0.000	0.002	0.042	0.009	0.000	0.054	0.012	0.005	0.101	0.006	0.000	0.001	0.045	0.000	0.111	0.000
Q'	0.004	0.000	0.005	0.000	0.001	0.003	0.000	0.000	0.000	0.001	0.011	0.003	0.003	0.143	0.006	0.000	0.017	0.002	0.111	0.000	0.020
Y'	0.000	0.001	0.053	0.000	0.007	0.000	0.001	0.008	0.000	0.025	0.003	0.022	0.000	0.008	0.000	0.002	0.074	0.006	0.000	0.020	0.000

Hereford

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.518	0.000	0.000	0.356	0.000	0.000	0.000	0.000	0.007	0.002	0.005	0.000	0.010	0.005	0.000	0.000	0.571	0.002	0.001	0.000
G_1	0.518	0.000	0.000	0.000	0.360	0.000	0.000	0.000	0.000	0.005	0.006	0.006	0.000	0.000	0.011	0.000	0.000	0.818	0.000	0.001	0.000
I_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
I_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K_1	0.356	0.360	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.004	0.002	0.000	0.000	0.001	0.000	0.000	0.397	0.000	0.000	0.000
O_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.003	0.035	0.008	0.024	0.000	0.015	0.000	0.000	0.000	0.000	0.009	0.000
P_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.002	0.000
Q_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.019	0.000
T_1	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.004	0.000
Y_2	0.007	0.005	0.001	0.000	0.007	0.003	0.000	0.000	0.000	0.000	0.008	0.157	0.000	0.001	0.126	0.000	0.000	0.002	0.000	0.044	0.002
A'_1	0.002	0.006	0.000	0.000	0.004	0.035	0.004	0.000	0.002	0.008	0.000	0.180	0.001	0.001	0.273	0.000	0.000	0.003	0.003	0.017	0.000
D'	0.005	0.006	0.001	0.000	0.002	0.008	0.004	0.000	0.000	0.157	0.180	0.000	0.004	0.001	0.487	0.000	0.000	0.006	0.012	0.071	0.002
F'_1	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.001	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
G'_1	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.001	0.000
I'_1	0.005	0.011	0.001	0.000	0.001	0.015	0.001	0.000	0.001	0.126	0.273	0.487	0.001	0.000	0.000	0.000	0.000	0.008	0.003	0.132	0.001
J'_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
O'_1	0.571	0.818	0.000	0.000	0.397	0.000	0.000	0.000	0.000	0.002	0.003	0.006	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.001	0.000
P'_1	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.012	0.000	0.009	0.003	0.000	0.000	0.000	0.000	0.000	0.025
Q'	0.001	0.001	0.000	0.000	0.000	0.009	0.002	0.019	0.004	0.044	0.017	0.071	0.000	0.001	0.132	0.000	0.000	0.001	0.000	0.000	0.000
Y'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.025	0.000	0.000

Holstein Friesian

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.001	0.006	0.000	0.007	0.076	0.002	0.005	0.000	0.001	0.001	0.002	0.002	0.003	0.003	0.002	0.002	0.003	0.000	0.008	0.002
G_1	0.001	0.000	0.001	0.001	0.005	0.000	0.000	0.001	0.000	0.010	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001
I_1	0.006	0.001	0.000	0.004	0.002	0.000	0.001	0.007	0.000	0.008	0.001	0.001	0.002	0.003	0.000	0.007	0.023	0.000	0.000	0.007	0.006
I_2	0.000	0.001	0.004	0.000	0.000	0.003	0.001	0.000	0.000	0.003	0.003	0.001	0.001	0.003	0.001	0.000	0.000	0.000	0.001	0.004	0.000
K_1	0.007	0.005	0.002	0.000	0.000	0.001	0.010	0.076	0.000	0.000	0.003	0.002	0.000	0.000	0.001	0.001	0.002	0.000	0.002	0.000	0.031
O_1	0.076	0.000	0.000	0.003	0.001	0.000	0.000	0.003	0.000	0.000	0.000	0.005	0.003	0.007	0.000	0.003	0.030	0.000	0.000	0.026	0.001
P_1	0.002	0.000	0.001	0.001	0.010	0.000	0.000	0.032	0.000	0.002	0.001	0.000	0.000	0.000	0.432	0.000	0.001	0.000	0.002	0.000	0.017
Q_2	0.005	0.001	0.007	0.000	0.076	0.003	0.032	0.000	0.000	0.000	0.005	0.002	0.000	0.001	0.011	0.001	0.009	0.001	0.006	0.000	0.063
T_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Y_2	0.001	0.010	0.008	0.003	0.000	0.000	0.002	0.000	0.000	0.000	0.013	0.000	0.005	0.016	0.007	0.002	0.005	0.010	0.003	0.010	0.000
A'_1	0.001	0.001	0.001	0.003	0.003	0.000	0.001	0.005	0.000	0.013	0.000	0.006	0.002	0.002	0.001	0.001	0.002	0.000	0.030	0.000	0.004
D'	0.002	0.000	0.001	0.001	0.002	0.005	0.000	0.002	0.000	0.000	0.006	0.000	0.158	0.114	0.001	0.003	0.001	0.094	0.001	0.011	0.001
F'_1	0.002	0.000	0.002	0.001	0.000	0.003	0.000	0.000	0.000	0.005	0.002	0.158	0.000	0.213	0.001	0.000	0.003	0.108	0.001	0.001	0.000
G'_1	0.003	0.000	0.003	0.003	0.000	0.007	0.000	0.001	0.000	0.016	0.002	0.114	0.213	0.000	0.000	0.000	0.006	0.147	0.092	0.022	0.001
I'_1	0.003	0.000	0.000	0.001	0.001	0.000	0.432	0.011	0.000	0.007	0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.001	0.014	0.000	0.013
J'_1	0.002	0.000	0.007	0.000	0.001	0.003	0.000	0.001	0.000	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.423	0.176	0.001	0.002	0.001
K'	0.002	0.000	0.023	0.000	0.002	0.030	0.001	0.009	0.000	0.005	0.002	0.001	0.003	0.006	0.001	0.423	0.000	0.055	0.000	0.008	0.008
O'_1	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.010	0.000	0.094	0.108	0.147	0.001	0.176	0.055	0.000	0.002	0.003	0.000
P'_1	0.000	0.001	0.000	0.001	0.002	0.000	0.002	0.006	0.000	0.003	0.030	0.001	0.001	0.092	0.014	0.001	0.000	0.002	0.000	0.038	0.004
Q'	0.008	0.001	0.007	0.004	0.000	0.026	0.000	0.000	0.000	0.010	0.000	0.011	0.001	0.022	0.000	0.002	0.008	0.003	0.038	0.000	0.000
Y'	0.002	0.001	0.006	0.000	0.031	0.001	0.017	0.063	0.000	0.000	0.004	0.001	0.000	0.001	0.013	0.001	0.008	0.000	0.004	0.000	0.000

Jersey

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.059	0.002	0.000	0.048	0.253	0.004	0.007	0.001	0.021	0.003	0.000	0.029	0.058	0.002	0.003	0.032	0.037	0.005	0.024	0.038
G_1	0.059	0.000	0.000	0.000	0.001	0.019	0.000	0.001	0.140	0.025	0.006	0.001	0.002	0.026	0.001	0.000	0.001	0.089	0.022	0.002	0.007
I_1	0.002	0.000	0.000	0.000	0.002	0.000	0.081	0.000	0.000	0.000	0.003	0.010	0.001	0.001	0.001	0.019	0.003	0.000	0.000	0.006	0.000
I_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K_1	0.048	0.001	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.168	0.011	0.051	0.007	0.014	0.004	0.001	0.040	0.010	0.000	0.006	0.015
O_1	0.253	0.019	0.000	0.000	0.000	0.000	0.002	0.037	0.000	0.003	0.014	0.000	0.000	0.070	0.008	0.004	0.195	0.014	0.000	0.036	0.013
P_1	0.004	0.000	0.081	0.000	0.000	0.002	0.000	0.000	0.000	0.001	0.009	0.000	0.000	0.000	0.002	0.000	0.002	0.001	0.000	0.001	0.006
Q_2	0.007	0.001	0.000	0.000	0.002	0.037	0.000	0.000	0.000	0.007	0.018	0.007	0.158	0.023	0.003	0.000	0.010	0.040	0.001	0.001	0.000
T_1	0.001	0.140	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.002	0.012	0.000	0.000	0.000
Y_2	0.021	0.025	0.000	0.000	0.168	0.003	0.001	0.007	0.001	0.000	0.001	0.004	0.000	0.002	0.016	0.000	0.045	0.021	0.003	0.024	0.073
A'_1	0.003	0.006	0.003	0.000	0.011	0.014	0.009	0.018	0.000	0.001	0.000	0.006	0.002	0.120	0.022	0.003	0.213	0.122	0.000	0.005	0.318
D'	0.000	0.001	0.010	0.000	0.051	0.000	0.000	0.007	0.000	0.004	0.006	0.000	0.026	0.023	0.011	0.002	0.001	0.010	0.002	0.004	0.008
F'_1	0.029	0.002	0.001	0.000	0.007	0.000	0.000	0.158	0.002	0.000	0.002	0.026	0.000	0.018	0.002	0.002	0.014	0.014	0.000	0.008	0.004
G'_1	0.058	0.026	0.001	0.000	0.014	0.070	0.000	0.023	0.000	0.002	0.120	0.023	0.018	0.000	0.002	0.008	0.072	0.317	0.060	0.069	0.285
I'_1	0.002	0.001	0.001	0.000	0.004	0.008	0.002	0.003	0.000	0.016	0.022	0.011	0.002	0.002	0.000	0.000	0.080	0.006	0.012	0.156	0.038
J'_1	0.003	0.000	0.019	0.000	0.001	0.004	0.000	0.000	0.000	0.000	0.003	0.002	0.002	0.008	0.000	0.000	0.009	0.010	0.000	0.012	0.003
K'	0.032	0.001	0.003	0.000	0.040	0.195	0.002	0.010	0.002	0.045	0.213	0.001	0.014	0.072	0.080	0.009	0.000	0.010	0.000	0.034	0.329
O'_1	0.037	0.089	0.000	0.000	0.010	0.014	0.001	0.040	0.012	0.021	0.122	0.010	0.014	0.317	0.006	0.010	0.010	0.000	0.055	0.055	0.076
P'_1	0.005	0.022	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.003	0.000	0.002	0.000	0.060	0.012	0.000	0.000	0.055	0.000	0.014	0.012
Q'	0.024	0.002	0.006	0.000	0.006	0.036	0.001	0.001	0.000	0.024	0.005	0.004	0.008	0.069	0.156	0.012	0.034	0.055	0.014	0.000	0.000
Y'	0.038	0.007	0.000	0.000	0.015	0.013	0.006	0.000	0.000	0.073	0.318	0.008	0.004	0.285	0.038	0.003	0.329	0.076	0.012	0.000	0.000

Limousin

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.024	0.001	0.008	0.007	0.032	0.001	0.001	0.000	0.000	0.006	0.022	0.007	0.011	0.003	0.006	0.000	0.016	0.012	0.016	0.000
G_1	0.024	0.000	0.001	0.003	0.000	0.000	0.000	0.001	0.000	0.001	0.002	0.000	0.001	0.000	0.000	0.008	0.001	0.018	0.002	0.001	0.000
I_1	0.001	0.001	0.000	0.007	0.000	0.007	0.001	0.005	0.000	0.000	0.029	0.015	0.009	0.005	0.020	0.005	0.006	0.009	0.014	0.000	0.001
I_2	0.008	0.003	0.007	0.000	0.000	0.000	0.001	0.005	0.000	0.000	0.000	0.003	0.004	0.003	0.002	0.214	0.005	0.018	0.000	0.010	0.001
K_1	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
O_1	0.032	0.000	0.007	0.000	0.000	0.000	0.000	0.005	0.002	0.003	0.000	0.009	0.010	0.000	0.000	0.012	0.160	0.000	0.002	0.006	0.000
P_1	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.045	0.000	0.000	0.001	0.056	0.002	0.000	0.001	0.000	0.002	0.000	0.299
Q_2	0.001	0.001	0.005	0.005	0.000	0.005	0.001	0.000	0.000	0.007	0.179	0.068	0.003	0.052	0.005	0.015	0.002	0.001	0.062	0.000	0.004
T_1	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.003	0.000	0.002	0.001	0.000
Y_2	0.000	0.001	0.000	0.000	0.001	0.003	0.045	0.007	0.000	0.000	0.019	0.003	0.003	0.001	0.002	0.000	0.002	0.016	0.000	0.002	0.046
A'_1	0.006	0.002	0.029	0.000	0.000	0.000	0.000	0.179	0.000	0.019	0.000	0.123	0.001	0.062	0.016	0.012	0.006	0.013	0.029	0.005	0.001
D'	0.022	0.000	0.015	0.003	0.000	0.009	0.000	0.068	0.000	0.003	0.123	0.000	0.018	0.004	0.029	0.015	0.002	0.007	0.001	0.012	0.000
F'_1	0.007	0.001	0.009	0.004	0.000	0.010	0.001	0.003	0.000	0.003	0.001	0.018	0.000	0.006	0.056	0.006	0.023	0.006	0.000	0.011	0.001
G'_1	0.011	0.000	0.005	0.003	0.000	0.000	0.056	0.052	0.000	0.001	0.062	0.004	0.006	0.000	0.012	0.001	0.000	0.028	0.222	0.006	0.028
I'_1	0.003	0.000	0.020	0.002	0.000	0.000	0.002	0.005	0.002	0.002	0.016	0.029	0.056	0.012	0.000	0.008	0.000	0.041	0.005	0.046	0.005
J'_1	0.006	0.008	0.005	0.214	0.000	0.012	0.000	0.015	0.000	0.000	0.012	0.015	0.006	0.001	0.008	0.000	0.000	0.025	0.002	0.001	0.000
K'	0.000	0.001	0.006	0.005	0.000	0.160	0.001	0.002	0.003	0.002	0.006	0.002	0.023	0.000	0.000	0.000	0.000	0.006	0.002	0.011	0.007
O'_1	0.016	0.018	0.009	0.018	0.000	0.000	0.000	0.001	0.000	0.016	0.013	0.007	0.006	0.028	0.041	0.025	0.006	0.000	0.000	0.001	0.002
P'_1	0.012	0.002	0.014	0.000	0.000	0.002	0.002	0.062	0.002	0.000	0.029	0.001	0.000	0.222	0.005	0.002	0.002	0.000	0.000	0.000	0.000
Q'	0.016	0.001	0.000	0.010	0.000	0.006	0.000	0.000	0.001	0.002	0.005	0.012	0.011	0.006	0.046	0.001	0.011	0.001	0.000	0.000	0.002
Y'	0.000	0.000	0.001	0.001	0.000	0.000	0.299	0.004	0.000	0.046	0.001	0.000	0.001	0.028	0.005	0.000	0.007	0.002	0.000	0.002	0.000

Poll Hereford

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.508	0.129	0.000	0.388	0.010	0.001	0.000	0.000	0.006	0.004	0.003	0.000	0.056	0.001	0.000	0.000	0.124	0.005	0.000	0.000
G_1	0.508	0.000	0.000	0.009	0.470	0.000	0.000	0.000	0.000	0.010	0.009	0.002	0.000	0.000	0.006	0.000	0.000	0.214	0.000	0.001	0.000
I_1	0.129	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
I_2	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000
K_1	0.388	0.470	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.009	0.003	0.000	0.000	0.002	0.000	0.000	0.191	0.000	0.000	0.000
O_1	0.010	0.000	0.011	0.000	0.000	0.000	0.000	0.003	0.000	0.005	0.002	0.000	0.032	0.046	0.001	0.000	0.000	0.012	0.036	0.008	0.000
P_1	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.000	0.013	0.012	0.000	0.000	0.000	0.001	0.000	0.000
Q_2	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.005	0.000	0.003	0.000	0.000	0.003	0.000	0.000	0.000	0.001	0.034	0.000
T_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Y_2	0.006	0.010	0.000	0.001	0.010	0.005	0.001	0.005	0.000	0.000	0.001	0.141	0.003	0.000	0.030	0.000	0.000	0.000	0.004	0.042	0.000
A'_1	0.004	0.009	0.000	0.000	0.009	0.002	0.002	0.000	0.000	0.001	0.000	0.075	0.000	0.000	0.146	0.000	0.000	0.004	0.002	0.022	0.000
D'	0.003	0.002	0.001	0.002	0.003	0.000	0.003	0.003	0.000	0.141	0.075	0.000	0.006	0.002	0.378	0.000	0.000	0.011	0.002	0.066	0.000
F'_1	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.003	0.000	0.006	0.000	0.000	0.014	0.000	0.000	0.368	0.000	0.000	0.000
G'_1	0.056	0.000	0.000	0.000	0.000	0.046	0.013	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.062	0.005	0.000
I'_1	0.001	0.006	0.001	0.003	0.002	0.001	0.012	0.003	0.000	0.030	0.146	0.378	0.014	0.001	0.000	0.000	0.000	0.028	0.000	0.114	0.000
J'_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
O'_1	0.124	0.214	0.000	0.000	0.191	0.012	0.000	0.000	0.000	0.000	0.004	0.011	0.368	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000
P'_1	0.005	0.000	0.000	0.000	0.000	0.036	0.001	0.001	0.000	0.004	0.002	0.002	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q'	0.000	0.001	0.000	0.000	0.000	0.008	0.000	0.034	0.000	0.042	0.022	0.066	0.000	0.005	0.114	0.000	0.000	0.000	0.000	0.000	0.000
Y'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Simmental

	B_1	G_1	I_1	I_2	K_1	O_1	P_1	Q_2	T_1	Y_2	A'_1	D'	F'_1	G'_1	I'_1	J'_1	K'	O'_1	P'_1	Q'	Y'
B_1	0.000	0.007	0.057	0.000	0.002	0.036	0.001	0.010	0.000	0.000	0.002	0.001	0.001	0.005	0.000	0.000	0.000	0.004	0.000	0.009	0.000
G_1	0.007	0.000	0.000	0.009	0.000	0.001	0.000	0.001	0.076	0.000	0.000	0.000	0.001	0.018	0.000	0.076	0.000	0.010	0.000	0.001	0.000
I_1	0.057	0.000	0.000	0.000	0.000	0.008	0.000	0.001	0.000	0.036	0.020	0.005	0.001	0.001	0.003	0.000	0.005	0.002	0.001	0.001	0.000
I_2	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.004	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
K_1	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
O_1	0.036	0.001	0.008	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.003	0.005	0.007	0.002	0.009	0.000	0.128	0.000	0.000	0.009	0.009
P_1	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.084	0.000	0.000	0.000	0.000	0.000	0.001
Q_2	0.010	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.010	0.002	0.000	0.002	0.000	0.000	0.004	0.000
T_1	0.000	0.076	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.014
Y_2	0.000	0.000	0.036	0.013	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.029	0.002	0.002	0.000	0.000	0.001	0.003	0.023	0.004	0.183
A'_1	0.002	0.000	0.020	0.004	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.044	0.000	0.007	0.003	0.000	0.000	0.048	0.001	0.008	0.004
D'	0.001	0.000	0.005	0.000	0.000	0.005	0.000	0.001	0.000	0.029	0.044	0.000	0.000	0.000	0.003	0.000	0.000	0.005	0.000	0.004	0.049
F'_1	0.001	0.001	0.001	0.003	0.000	0.007	0.000	0.000	0.031	0.002	0.000	0.000	0.000	0.004	0.001	0.031	0.232	0.000	0.004	0.000	0.001
G'_1	0.005	0.018	0.001	0.000	0.003	0.002	0.000	0.010	0.000	0.002	0.007	0.000	0.004	0.000	0.005	0.000	0.003	0.170	0.022	0.010	0.005
I'_1	0.000	0.000	0.003	0.000	0.000	0.009	0.084	0.002	0.000	0.000	0.003	0.003	0.001	0.005	0.000	0.000	0.001	0.002	0.000	0.025	0.002
J'_1	0.000	0.076	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.014
K'	0.000	0.000	0.005	0.000	0.000	0.128	0.000	0.002	0.010	0.001	0.000	0.000	0.232	0.003	0.001	0.010	0.000	0.000	0.000	0.003	0.000
O'_1	0.004	0.010	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.048	0.005	0.000	0.170	0.002	0.000	0.000	0.000	0.002	0.002	0.003
P'_1	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.001	0.000	0.004	0.022	0.000	0.000	0.000	0.002	0.000	0.031	0.019
Q'	0.009	0.001	0.001	0.000	0.000	0.009	0.000	0.004	0.000	0.004	0.008	0.004	0.000	0.010	0.025	0.000	0.003	0.002	0.031	0.000	0.000
Y'	0.000	0.000	0.000	0.001	0.000	0.009	0.001	0.000	0.014	0.183	0.004	0.049	0.001	0.005	0.002	0.014	0.000	0.003	0.019	0.000	0.000

Appendix IV

Population sizes of cattle breeds in Great Britain

(recorded by the Rare Breeds Survival Trust for the Ministry of Agriculture, Fisheries and Food 1996)

Breed	No. males	No. females	N_e[1]
Aberdeen Angus	500	10,000	1,905
Ayrshire	1,200	120,000	4,752
Belgian Blue	102	735	358
Belted Galloway	239	2,150	860
Blonde d'Aquitaine	3,806	14,026	11,975
British White	102	1,246	377
Charolais	3,000	12,000	9,600
Chillingham	15	21	35
Devon	100	2,000	381
Dexter	121	1,800	454
Galloway	200	6,000	774
Gloucester	34	500	127
Guernsey	100	22,000	398
Hereford	1,820	7,200	5,811
Hereford (traditional)	27	364	101
Highland	41	2,000	161
Holstein-Friesian	6,189	175,000	23,910
Irish Moiled	18	194	66
Jersey	500	8,000	1,882
Kerry	11	97	40
Limousin	4,392	12,000	12,861
Lincoln Red	73	1,758	280
Longhorn	250	1,700	872
Murray Grey	35	1,050	135
Red Poll	34	957	131
Shetland	31	358	114
Beef Shorthorn	34	549	128
Dairy Shorthorn	270	4,000	1,012
Simmental	4,054	4,582	8,604
South Devon	589	7,100	2,176
Sussex	300	7,000	1,150
Welsh Black	103	2,100	393
White Park	31	367	114
Whitebred Shorthorn	150	119	265

[1] Estimated from the numbers of males (Nm) and females (Nf) where Ne = 4NmNf/Nm+Nf (Falconer 1989)

Bibliography

Adalsteinsson, S., Bjarnadottir, S., Vage, D.I. and Jonmundsson, J.V. (1995) Brown coat colour in Icelandic cattle produced by the loci *Extension* and *Agouti*. *Journal of Heredity* **86**, 395-398.

Ajmone-Marsan, P., Kuiper, M., Valentini, A., Cassandro, M. and Bertoni, G. (1996) Evaluation of genetic diversity in Friesian cattle with the aid of AFLP markers. *Proceedings 25th International Conference on Animal Genetics, Tours, France.*

Anderson, T.W. (1984) *An introduction to multivariate statistical analysis*. Second edition. Wiley, New York, USA.

Andersson-Eklund, L. (1993) *Genetic markers and quantitative traits in dairy cattle*. PhD thesis. Swedish University of Agricultural Sciences, Uppsala, Sweden.

Astolfi, P., Pagnacco, G. and Guglielmino-Matessi, C.R. (1983) Phylogenetic analysis of native Italian cattle breeds. *Journal of Animal Breeding and Genetics* **100**, 87-100.

Avise, J.C. (1989) Gene trees and organismal histories - a phylogenetic approach to population biology. *Evolution* **43**, 1192-1208.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E. (1987) Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics. *Annual Review of Ecological Systematics* **18**, 489-522.

Baker, C.M.A. and Manwell, C. (1980) Chemical classification of cattle. 1. Breed groups. *Animal Blood Groups and Biochemical Genetics* **11**, 127-150.

Baker, C.M.A. and Manwell, C. (1991) Population genetics, molecular markers and gene conservation of bovine breeds. In: *Cattle Genetic Resources* (ed. by C.G. Hickman), pp. 221-304. Elsevier, Amsterdam.

Barendse, W., Vaiman, D., Kemp, S.J., Sugimoto, Y. et al. (1997) A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* **8**, 21-28.

Baret, P.V. and Hill, W.G. (1997) Gametic disequilibrium mapping: potential applications in livestock. *Animal Breeding Abstracts* **65**, 309-318.

Barker, J.S.F., Bradley, D.G., Fries, R., Hill, W.G., Nei, M. and Wayne, R.K. (1993) An integrated global programme to establish the genetic relationships among the breeds of each domestic animal species. *Report of a Working Group for Animal Production and Health Division, Food and Agriculture Organization of the United Nations (FAO)*, Rome.

Barton, N.H. and Slatkin, M. (1986) A quasi-equilibrium theory of the distribution of rare alleles in a sub-divided population. *Heredity* **56**, 409-415.

Basedow, M., Barre-Dirie, Harlizius, B., Looft, C., Simon, D. and Kalm, E. (1996) The estimation of genetic distance between German cattle breeds. *47th Annual Meeting of the European Association of Animal Production (EAAP)*, Lillehammer, Norway.

Beavis, W.D., Grant, D., Albertsen, M. and Fincher, R. (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative loci. *Theoretical and Applied Genetics* **83**, 141-145.

Boichard, D., Maignel, L. and Verrier, E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* **29**, 5-23.

Bouw, J., Buys, C. and Schreuder, I. (1974) Further studies on the genetic control of the blood group system C of cattle. *Animal Blood Groups and Biochemical Genetics* **5**, 105-114.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-457.

Bradley, D.G., MacHugh, D.E., Loftus, R.T., Sow, R.S., Hoste, C.H. and Cunningham ,E.P. (1994) Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Animal Genetics* **25**, 7-12.

Brenneman, R.A., Davis, S.K., Sanders, J.O., Burns, B.M., Wheeler, T.C., Turner, J.W. and Taylor, J.F. (1996) The polled locus maps to BTA1 in a *Bos indicus* x *Bos taurus* cross. *Journal of Heredity* **87**, 156-161.

Brown, J.K.M. (1994) Bootstrap hypothesis tests for evolutionary trees and other dendrograms. *Proceedings of the National Academy of Sciences* **91**, 12293-12297.

Buchanan, F.C., Adams, L.J., Littlejohn, R.P., Maddox, J.F. and Crawford, A.M. (1994) Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* **22**, 397-403.

Buis, R.C. and Tucker, E.M. (1983) Relationships between rare breeds of sheep in the Netherlands as based on blood typing. *Animal Blood Groups and Biochemical Genetics* **14**, 17-26.

Buys, C. and Chiperzak, J. (1992) A comparative study of blood groups in the Kerry and Dexter cattle breeds. In: *Genetic Conservation of Domestic Livestock* (ed. by L. Alderson and I. Bodo), CAB International, Wallingford, UK.

Cabellero, A. (1994) Developments in the prediction of effective population size. *Heredity* **73**, 657-679.

Cabellero, A. and Hill, W.G. (1992) A note on the inbreeding effective population size. *Evolution* **46**, 1969-1972.

Cabellero, A. and Hill, W.G. (1992) Effective size of non-random mating populations. *Genetics* **130**, 909-916.

Cavalli-Sforza, L.L. and Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 550-570.

Ceppellini, R., Siniscalca, M. and Smith, C.A.B. (1955) The estimation of gene frequencies in a random mating population. *Annals of Human Genetics* **20**, 97-117.

Chakraborty, R. (1985) Genetic distance and gene diversity: some statistical considerations. In: *Multivariate Analysis* (ed. by P. Krishnaiah), pp 77-96, Elsevier, Amsterdam.

Chakraborty, R., Kamboh, M.I., Nwankwo, M. and Ferrell, R.E. (1992) Caucasian genes in American blacks: new data. *American Journal of Human Genetics* **50**, 145-155.

Chakraborty, R., Srinivasan, M.R. and Daiger, S.P. (1993) Evaluation of standard error and confidence interval of estimated multilocus genotype probabilities and their implications in DNA forensics. *American Journal of Human Genetics* **52**, 60-70.

Charlier, C., Denys, B., Belanche, J.I., Coppieters, W., Grobet, L., Mni, M., Womack, J., Hanset, R. and Georges, M. (1996) Microsatellite mapping of the bovine *roan* locus: a major determinant of White Heifer disease. *Mammalian Genome* **7**, 138-142.

Chesser, R.K. (1991) Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics* **129**, 573-583.

Chesser, R.K., Rhodes, O.E., Sugg, D.W. and Schnabel, A. (1993) Effective sizes for subdivided populations. *Genetics* **135**, 1221-1232.

Chiano, M.N. and Yates, J.R.W. (1994) Bootstrapping in human genetic linkage. *Annals of Human Genetics* **58**, 129-143.

Ciampolini, R., Moazami-Goudarzi, K., Vaiman, D., Dillmann, C., Mazzanti, E., Foulley, J-L., Leveziel, H. and Cianci, D. (1995) Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *Journal of Animal Science* **73**, 3259-3268.

Crandall, K.A. and Templeton, A.R. (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**, 959-969.

Cuadras, C.M. (1989) Distance analysis in discrimination and classification using both continuous and categorical variables. In: *Statistical Data Analysis and Inference* (ed. Y. Dodge), pp. 459-473. North-Holland, Amsterdam, The Netherlands.

Cuadras, C.M. (1992) Some examples of distance based discrimination. *Biometrical Letters* **29**, 3-20.

Cunningham, E.P., Loftus, R.T., MacHugh, D.E. and Bradley, D.G. (1994) Molecular evolution of African, European and Asian cattle. *Proceedings 5th World Congress on Genetics Applied to Livestock Production* **21**, 86-89.

Davidson, W.S., Birt, T.P. and Green, J.M. (1989) A review of genetic variation in Atlantic salmon, *Salmo salar L.*, and its importance for stock identification, enhancement programmes and aquaculture. *Journal of Fish Biology* **34**, 547-560.

Delahunty, C., Ankener, W., Deng, Q., Eng, J. and Nickerson, D. (1996) Testing the feasibility of DNA typing for human identification by PCR and oligonucleotide ligation assay. *American Journal of Human Genetics* **58**, 1239-1246.

Dillmann, C., BarHen, A., Guerin, D., Charcosset, A. and Murigneux, A. (1997) Comparison of RFLP and morphological distances between maize *Zea mays L* inbred lines. Consequences for germplasm protection purposes. *Theoretical and Applied Genetics* **95**, 92-102.

Dunnington, E.A., Gal, O., Siegel, P.B., Haberfeld, A., Cahaner, A., Lavi, U., Plotsky, Y. and Hillel, J. (1991) Deoxyribonucleic acid fingerprint comparisons between selected populations of chickens. *Poultry Science* **70**, 463-467.

Dunnington, E.A., Stallard, L.C., Hillel, J. and Siegel, P.B. (1994) Genetic diversity among commercial chicken populations estimated from DNA fingerprints. *Poultry Science* **73**, 1218-1225.

EAAP Working Group on Animal Genetic Resources. (1993) *Genetic diversity of European livestock breeds*. (ed. by D.L. Simon and D. Buchenauer) EAAP Publication No. 66, Wageningen.

Efron, B. and Tibshirani, R.J. (1993) *An introduction to the bootstrap*. Chapman and Hall, New York, London.

Efron, B. and Tibshirani, R.J. (1986) Bootstrap measures for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science* **1**, 54-77.

Eggen, A. and Fries, R. (1995) An integrated cytogenetic and meiotic map of the bovine genome. *Animal Genetics* **26**, 215-236.

Evett, I.W., Pinchin, R. and Buffery, C. (1992) An investigation of the feasibility of inferring ethnic origin from DNA profiles. *Journal of the Forensic Science Society* **32**, 301-306.

Falconer, D.A. (1989) *Introduction to Quantitative Genetics*, third edition. Longman , London.

Felsenstein, J. (1976) The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* **10**, 253-280.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Felsenstein, J. (1995) PHYLIP (Phylogenetic Inference Package) version 3.572c. Department of Genetics, University of Washington, USA.

Food and Agriculture Organization of the United Nations (FAO). (1995) *World Watch List for Domestic Animal Diversity.* Second edition (ed. B. Scherf), Rome.

Food and Agriculture Organization of the United Nations (FAO). (1996) *Global Project for the Maintenance of Domestic Animal Genetic Diversity* (MoDAD) http://www.fao.org/dad-is/

Galvin, P., McKinnell, S., Taggart, J.B., Ferguson, A., O'Farrell, M. and Cross, T.F. (1995) Genetic stock identification of Atlantic salmon using single locus minisatellite DNA profiles. *Journal of Fish Biology* **47**, 186-199.

Genstat 5.3. (1993) Release 5.3 Reference Manual. Clarendon Press, Oxford.

Georges, M. and Andersson, L. (1996) Livestock genomics comes of age. *Genome Research* **6**, 907-921.

Georges, M., Drinkwater, R., King, T., Mishra, A., Moore, S.S., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X.Y., Womack, J.E. and Hetzel, J. (1993) Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nature Genetics* **4**, 206-210.

Goddard, M.E. (1992) Optimal effective population size for the global population of black-and-white dairy cattle. *Journal of Dairy Science* **75**, 2902-2911.

Goldstein, D.B. and Pollock, D.D. (1994) Least squares estimation of molecular distance - noise abatement in phylogenetic reconstruction. *Theoretical Population Biology* **45**, 219-226.

Goldstein, D.B. and Pollock, D.D. (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335-342.

Goldstein, D.B., Ruiz-Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Gonzalez, P., Tunon, M.J. and Vallejo, M. (1987) Genetic relationship between seven Spanish native breeds of cattle. *Animal Genetics* **18**, 249-256.

Grobet, L., Royo Martin, J.L., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., Fries, R., Hanset, R. and Georges, M. (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 71-74.

Grosclaude, F., Aupetit, R.Y., Lefebvre, J. and Meriaux, J.C. (1990) Tentative analysis of genetic relationships between French cattle breeds using biochemical polymorphism. *Genetics Selection Evolution* **22**, 317-338.

Grosclaude, F., Guerin, G. and Houlier, G. (1979) The genetic map of the B system of cattle blood groups as observed in French breeds. *Animal Blood Groups and Biochemical Genetics* **10**, 199-218.

Grosclaude, F., Lefebvre, J. and Noe, G. (1983) Additional data on the genetic map of the B system of cattle blood groups. *Genetics Selection Evolution* **15**, 45-54.

Guerin, G., Grosclaude, F. and Houlier, G. (1981) The C system of cattle blood groups. 2. Partial genetic map of the system. *Animal Blood Groups and Biochemical Genetics* **12**, 15-21.

Gwakisa, P.S., Kemp, S.J. and Teale, A.J. (1994) Characterization of Zebu cattle breeds in Tanzania using random amplified polymorphic DNA markers. *Animal Genetics* **25**, 89-94.

Hall, S.J.G. and Bradley, D.G. (1995) Conserving livestock breed biodiversity. *Trends in Ecology and Evolution* **10**, 267-270.

Hartl, D.L. (1981) *A primer of population genetics*. Sinauer, Sunderland, Massachusetts USA.

Heath-Agnew, E. (1983) *A history of Hereford cattle and their breeders*. Duckworth, London.

Hermans, I.F., Morris, C.A., Chambers, G.K., Towers, N.R. and Jordan, T.W. (1993) Assessment of DNA fingerprinting for determining genetic diversity in sheep populations. *Animal Genetics* **24**, 385-388.

Hill, W.G. (1985) Effects of population size on response to short and long term selection. *Journal of Animal Breeding and Genetics* **102**, 161-173.

Hill, W.G. (1986) Population size and design of breeding programmes. *Proceedings 3rd World Congress of Genetics Applied to Livestock Production* **12**, 245-256.

Hill, W.G. (1974) Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229-239.

Hillis, D.M. (1984) Misuse and modification of Nei's genetic distance. *Systematic Zoology* **33**, 238-240.

Hillis, D.M. and Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**, 182-192.

Hongtrakul, V., Huestis, G.M. and Knapp, S.J. (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. *Theoretical and Applied Genetics* **95**, 400-407.

Hudson, R.R., Boos, D.D. and Kaplan, N.L. (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138-151.

Jackson, I.J. (1994) Molecular and developmental genetics of mouse coat colour. *Annual Review of Genetics* **28**, 189-217.

Jacquard, A. (1974) *The Genetic Structure of Populations*. Springer-Verlag, New York.

Joerg, H., Fries, H.R., Meijerink, E. and Stranzinger, G.F. (1996) Red coat colour in Holstein cattle is associated with a deletion in the MSHR gene. *Mammalian Genome* **7**, 317-318.

Jorde, L.B., Watkins, W.S., Carlson, M., Groden, J., Albertsen, H., Thliveris, A. and Leppert, M. (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *American Journal of Human Genetics* **54**, 884-898.

Kidd, K.K. and Pirchner, F. (1971) Genetic relationships of Austrian cattle breeds. *Animal Blood Groups and Biochemical Genetics* **2**, 145-158.

Kidd, K.K., Stone, W.H., Crimella, C., Carenzi, C., Casati, M. and Rognoni, G. (1980) Immunogenetic and population genetic analyses of Iberian cattle. *Animal Blood Groups and Biochemical Genetics* **11**, 21-38.

Kimura, M. (1983) *The neutral theory of evolution*. Cambridge University Press.

Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S. and Lien, S. (1995) The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat colour determination. *Mammalian Genome* **6**, 636-639.

Koots, K.R. and Crow, G.H. (1989) Breed structure of pedigreed Hereford cattle in Canada. *Canadian Journal of Animal Science* **69**, 627-634.

Krzanowski, W.J. (1988) *Principles of multivariate analysis: a user's perspective*. Clarendon Press, Oxford, England.

Krzanowski, W.J. and Marriott, F.H.C. (1995) *Multivariate analysis. Part II. Classification, covariance structures and repeated measurements*. Kendall's Library of Statistics. Edward Arnold, London.

Kwok, P-Y., Carlson, C., Yager, T., Ankener, W. and Nickerson, D. (1994) Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* **23**, 138-149.

Lagziel, A., Lipkin, E. and Soller, M. (1996) Association between SSCP haplotypes at the bovine growth hormone gene and milk protein percentage. *Genetics* **142**, 945-951.

Langley, C.H., Tobari, Y.N. and Kojima, K. (1974) Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**, 921-936.

Li, W-H. and Graur, D. (1991) *Fundamentals of molecular evolution*. Sinauer, Sunderland, Massachussets, USA.

Lidauer, M. and Mantysaari, E. (1996) Genetic constitution of the Finnish black and white cattle population and the influence of Holsteinization on protein yield, days open and somatic cell count. *Acta Agriculturae Scandinavica* **46**, 193-200.

Litt, M. and Jorde, L.B. (1986) Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2*q*. *American Journal of Human Genetics* **39**, 166-178.

Liu, R.Y. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In: *Exploring the limits of bootstrap* (ed. LePage and Billard). John Wiley, New York.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M. and Cunningham, E.P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences* **91**, 2757-2761.

Long, J.C. (1991) The genetic structure of admixed populations. *Genetics* **127**, 417-428.

Long, J.C. and Smouse, P.E. (1983) Intertribal gene flow between the Yecuana and Yanomama: genetic analysis of an admixed village. *American Journal of Physical Anthropology* **61**, 411-422.

Lu, J., Knox, M.R., Ambrose, M.J., Brown, J.K.M and Ellis, T.H.N. (1996) Comparative analysis of genetic diversity in pea assessed by RFLP-based and PCR-based methods. *Theoretical and Applied Genetics* **93**, 1103-1111.

MacHugh, D.E. (1996) *Molecular biogeography and genetic structure of domesticated cattle*. Ph.D thesis. University of Dublin, Ireland.

MacHugh, D.E., Loftus, R.T., Bradley, D.G., Sharp, P.M. and Cunningham, E.P. (1994) Microsatellite DNA variation within and among European cattle breeds. *Proceedings of the Royal Society of London Part B* **256**, 25-31.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, E.P. and Bradley, D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**, 1071-1086.

Maddison, W.P. (1996) Molecular approaches and the growth of phylogenetic biology. In: *Molecular Zoology. Advances, strategies and protocols* (ed. J.D. Ferraris and S.R. Palumbi) pp 47-61,Wiley-Liss, New York.

Manly, B.F.J. (1986) *Multivariate statistical methods: A primer*. Chapman and Hall, London.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.

Manwell, C. and Baker, C.M.A. (1980) Chemical classification of cattle. 2. Phylogenetic tree and specific status of Zebu. *Animal Blood Groups and Biochemical Genetics* **11**, 151-162.

Matusita, K. (1956) Decision rule, based on distance, for the classification problem. *Annals of the Institute of Statistical Mathematics* **8**, 67-77.

May, R.M. (1990) Taxonomy as destiny. *Nature* **347**, 129-130.

McDade, L. (1992) Hybrids and phylogenetic systematics. II. The impact of hybrids on cladistic analysis. *Evolution* **46**, 1329-1346.

Medjugorac, I., Kustermann, W., Lazar, P., Russ, I. and Pirchner, F. (1994) Marker derived phylogeny of European cattle supports demic expansion of agriculture. *Animal Genetics* **25**, 19-27.

Mejdell, C.M., Lingaas, F., Stear, M.J. and Lie, O. (1993) Genetic analysis of Major Histocompatibility Complex Class I antigens, serum transferrins and red blood cell antigens in Norwegian breeds of cattle. *Acta Agriculturae Scandinavica* **43**, 193-200.

Meuwissen, T.H.E. and Woolliams, J.A. (1994) Effective sizes of livestock populations to prevent a decline in fitness. *Theoretical and Applied Genetics* **89**, 1019-1026.

Miglior, F. and Burnside, E.B. (1995) Inbreeding of Canadian Holstein cattle. *Journal of Dairy Science* **78**, 1163-1167.

Miglior, F., Burnside, E.B. and Dekkers, J.C.M. (1995) Non-additive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. *Journal of Dairy Science* **78**, 1168-1173.

Miglior, F., Burnside, E.B. and Kennedy, B.W. (1995) Production traits of Holstein cattle - estimation of non-additive genetic variance components and inbreeding depression. *Journal of Dairy Science* **78**, 1174-1180.

Miglior, F., Szkotnicki, B. and Burnside, E.B. (1992) Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. *Journal of Dairy Science* **75**, 1112-1118.

Miller, K.M., Withler, R.E. and Beacham, T.D. (1996) Stock identification of *coho* salmon (*Oncorhynchus kisutch*) using minisatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Science* **53**, 181-195.

Ministry of Agriculture, Fisheries and Food (MAFF). (1996) *UK Farm Animal Genetic Resources*. Minutes of the 2nd Annual Meeting of Interested Parties.

Mirck, M.H., von Bannisseht Wijsmuller, T., Timmermans Besselink, W.J.H., Van Luijk, J.H.L. and Bunter, J.B. (1995) Optimization of the PCR test for the mutation causing bovine leukocyte adhesion deficiency. *Cellular and Molecular Biology* **41**, 695-698.

Moazami-Goudarzi, K., Vaiman, D., Mercier, D., Grohs, C., Furet, J.P., Leveziel, H. and Martin, P. (1994) Analysis of genetic diversity in French cattle breeds by the use of microsatellites: preliminary results. *Genetics Selection Evolution* **26**, 155-165.

Moritz, C., Dowling, T.E. and Brown, W.M. (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**, 269-292.

Mueller, L.D. (1979) A comparison of two methods for making statistical inferences on Nei's measure of genetic distance. *Biometrics* **35**, 757-763.

Nei, M. (1972) Genetic distance between populations. *American Naturalist* **106**, 282-292.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323.

Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **41**, 225-233.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.

Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press.

Nei, M. and Chakravarti, A. (1977) Drift variances of Fst and Gst statistics obtained from a finite number of isolated populations. *Theoretical Population Biology* **11**, 307-325.

Nei, M. and Chesser, R.K. (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253-259.

Nei, M. and Roychoudhury, A.K. (1974) Sampling variances of heterozygosity and genetic distance. *Genetics* **76**, 379-390.

Nei, M., Tajima, F. and Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153-170.

Neimann-Sorensen, A. (1956) Blood groups and breed structure as exemplified by three Danish breeds. *Acta Agriculturae Scandinavica* **6**, 115-137.

Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. and Landegren, U. (1990) Automated DNA diagnostics using an ELISA-based oligonucleotide assay. *Proceedings of the National Academy of Sciences* **87**, 8923-8927.

Nikiforov, T.T., Rendle, R.B., Goelet, P., Rogers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L. and Knapp, M.R. (1994) Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Research* **22**, 4167-4175.

O'hUigin, C. and Cunningham, E.P. (1990) Analysis of breeding structure of the Kerry breed. *Journal of Animal Breeding and Genetics* **107**, 452-457.

Ota, T. DISPAN. A program for computing Nei's standard genetic distance and its standard error. http//www.bio.psu.edu/imeg/neilab.htm

Ott, J. (1991) *Analysis of Human Genetic Linkage*. John Hopkins University Press, Baltimore, USA.

Ozkutuk, K. and Bichard, M. (1977) Studies of pedigree Hereford cattle breeding 1. Herdbook analyses. *Animal Production* **24**, 1-13.

Paterson, A.H. (1995) Molecular dissection of quantitative traits: progress and prospects. *Genome Research* **5**, 321-333.

Paterson, A.H., Lin, Y-R., Zhikang, L., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S-C., Stansel, J.W. and Irvine, J.E. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714-1718.

Paul, S., Wachira, F.N., Powell, W. and Waugh, R. (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis (L) O. Kuntze*) revealed by AFLP markers. *Theoretical and Applied Genetics* **94**, 255-263.

Perez-Enciso, M., Fernando, R.L., and Gianola, D. (1992) The use of the uncertain relationship matrix for computing inbreeding effective population size. *Proceedings 43rd Meeting of the European Association of Animal Production (EAAP)*, p 264.

Porter, V. (1991) *Cattle. A handbook to the breeds of the world.* Christopher Helm, London.

Rao, C.R. (1973) *Linear statistical inference and its applications.* Second edition. Wiley, New York, USA.

Raymond, M. and Rousset, F. (1995) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Rendel, J. (1967) Studies of blood groups and protein variants as a means of revealing similarities and differences between animal populations. *Animal Breeding Abstracts* **35**, 371-383.

Reynolds, J., Weir, B.S. and Cockerham, C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**, 767-779.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Robertson, A. and Asker, A.A. (1951) The genetic history and breed structure of British Friesian cattle. *Empire Journal of Experimental Agriculture* **19**, 113-130.

Roff, D.A. and Bentzen, P. (1989) The statistical analysis of mitochondrial DNA polymorphisms: chi-squared and the problem of small samples. *Molecular Biology and Evolution* **6**, 539-545.

Rogers, J.S. (1972) Measures of genetic similarity and genetic distance. *Studies in Genetics VII.* University of Texas publication 7213, 145-153.

Royle, N.J. (1983) *Polymorphisms in rare breeds of cattle.* Ph.D thesis. University of Reading, England.

Ruiterkamp, W.A., Spek, C.M. and Bouw, J. (1977) Gene clusters in the blood group system B of cattle. *Animal Blood Groups and Biochemical Genetics* **8**, 231-240.

Saitou, N. and Nei, M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Sanchez, A., Ocana, J. and Utzet, F. (1995) Sampling theory, estimation and significance testing for Prevosti's estimate of genetic distance. *Biometrics* **51**, 1216-1235.

Santiago, E. and Caballero, A. (1995) Effective size of populations under selection. *Genetics* **139**, 1013-1030.

Schwenger, B., Schober, S. and Simon, D. (1993) DUMPS cattle carry a point mutation in the uridine monophosphate synthase gene. *Genomics* **16**, 241-244.

Searle, A.G. (1968) *Comparative genetics of coat colour in mammals*. Logos Press, London.

Shaklee, J.B. and Varnavskaya, N.V. (1994) Electrophoretic characterization of odd year pink salmon (*Oncorhynchus gorbuscha*) populations from the Pacific coast of Russia and comparison with selected North American populations. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 158-171.

Shriver, M.D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R.E. and Chakraborty, R. (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution* **12**, 914-920.

Siegel, P.B., Haberfeld, A., Mukherjee, T.K., Stallard, L.C., Marks, H.L., Anthony, N.B. and Dunnington, E.A. (1992) Jungle fowl domestic fowl relationships - a use of DNA fingerprinting. *World's Poultry Science Journal* **48**, 147-155.

Sigurdsson, A. and Jonmundsson, J.V. (1995) Inbreeding and its impact in the closed population of Icelandic dairy cattle. *Acta Agriculturae Scandinavica* **45**, 11-16.

Slatkin, M. (1985) Rare alleles as indicators of gene flow. *Evolution* **39**, 53-65.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.

Slatkin, M. (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* **64**, 71-74.

Slatkin, M. and Arter, H.E. (1991) Spatial autocorrelation methods in population genetics. *American Naturalist* **138**, 499-517.

Slatkin, M. and Maddison, W.P. (1990) Detecting isolation by distance using phylogenies of genes. *Genetics* **126**, 249-260.

Smith, C.A.B. (1977) A note on genetic distance. *Annals of Human Genetics* **40**, 463-479.

Smith, E.J., Jones, C.P., Bartlett, J. and Nestor, K.E. (1996) Use of randomly amplified polymorphic DNA markers for the genetic analysis of relatedness and diversity in chickens and turkeys. *Poultry Science* **75**, 579-584.

Sneath, R.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*. Freeman, San Francisco, USA.

Stormont, C. (1955) Linked genes, pseudoalleles and blood groups. *American Naturalist* **89**, 105-116.

Swofford, D.L. and Olsen, G.J. (1990) Phylogeny reconstruction. In: *Molecular Systematics* (ed. D.M. Hillis and C. Moritz). Sinauer, Sunderland, Massachusetts, USA.

Syvanen, A-C., Ikonen, E., Manninen, T., Bengstrom, M., Soderlund, H., Aula, P. and Peltonen, L. (1992) Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: application to aspartylglucosaminuria in Finland. *Genomics* **12**, 590-595.

Tajima, F. and Takezaki, N. (1994) Estimation of evolutionary distance for reconstructing molecular phylogenetic trees. *Molecular Biology and Evolution* **11**, 278-286.

Takezaki, N. and Nei, M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389-399.

Te Braake, M.F.H., Groen, A.F. and Van der Lugt, A.W. (1994) Trends in inbreeding in Dutch black-and-white dairy cattle. *Journal of Animal Breeding and Genetics* **111**, 356-366.

Templeton, A.R., Boerwinkle, E. and Sing, C.F. (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. 1. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. *Genetics* **117**, 343-351.

Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., Krings, M., Paabo, S., Watson, E., Risch, N., Jenkins, T. and Kidd, K.K. (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380-1387.

Tunon, M.J., Gonzalez, P. and Vallejo, M. (1989) Genetic relationships between fourteen native Spanish breeds of goat. *Animal Genetics* **20**, 205-212.

Van Dongen, S. (1995) How should we bootstrap allozyme data? *Heredity* **74**, 445-447.

Van Dongen, S. and Backeljau, T. (1995) One- and two- sample tests for single locus inbreeding coefficients using the bootstrap. *Heredity* **74**, 129-135.

Van Doornik, D.M., Milner, G.B. and Winans, G.A. (1996) Transferrin polymorphism in coho salmon (*Oncorynchus kisutch*) and application to genetic stock identification. *Fishery Bulletin* **94**, 566-575.

Van Zeveren, A., Bouquet, Y., Van de Weghe, A. and Coppieters, W. (1990) A genetic blood marker study on four pig breeds I. Estimation and comparison of within-breed variation. *Journal of Animal Breeding and Genetics* **107**, 104-112.

Van Zeveren, A., Bouquet, Y., Van de Weghe, A. and Coppieters, W. (1990) A genetic blood marker study on four pig breeds. 2. Genetic relationship between the populations. *Journal of Animal Breeding and Genetics* **107**, 113-118.

Villanueva, B. and Woolliams, J.A. (1997) Optimization of breeding programmes under index selection and constrained inbreeding. *Genetical Research* **69**, 145-158.

Visscher, P.M., Thompson, R. and Haley, C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013-1020.

Waples, R.S. (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-391.

Weir, B.S. (1996) *Genetic Data Analysis II.* Sinauer, Sunderland, Massachusetts, USA.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Welch, B.L. (1939) Note on discriminant functions. *Biometrika* **31**, 218-220.

Williams, R.C., Knowler, W.C., Pettitt, D.J., Long, J.C., Rokala, D.A., Polesky, H.F., Hackenberg, R.A., Steinberg, A.G. and Bennett, P.H. (1992) The magnitude and origin of European-American admixture in the Gila River Indian community of Arizona: a union of genetics and demography. *American Journal of Human Genetics* **51**, 101-110.

Woolliams, J.A. and Mantysaari, E.A. (1995) Genetic contributions of Finnish Ayrshire bulls over four generations. *Animal Science* **61**, 177-187.

Woolliams, J.A., Wray, N.R. and Thompson, R. (1993) Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genetical Research* **62**, 231-242.

Workman, P.L. and Niswander, J.D. (1970) Population studies on Southwestern Indian tribes II. Local genetic differentiation in the Papago. *American Journal of Human Genetics* **22**, 24-49.

Wray, N.R. and Thompson, R. (1990) Predictions of rates of inbreeding in selected populations. *Genetical Research* **55**, 41-54.

Wray, N.R., Woolliams, J.A. and Thompson, R. (1990) Methods for predicting rates of inbreeding in selected populations. *Theoretical and Applied Genetics* **80**, 503-512.

Wray, N.R., Woolliams, J.A. and Thompson, R. (1994) Prediction of rates of inbreeding in populations undergoing index selection. *Theoretical and Applied Genetics* **87**, 878-892.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Wright, S. (1978) *Variability Within and Among Natural Populations*. Evolution and the Genetics of Populations, Volume 4. University of Chicago Press, Chicago, Illinois, USA.

Zanotti Casati, M., Gandini, G.C. and Leone, P. (1990) Genetic variation and distances of five Italian native sheep breeds. *Animal Genetics* **21**, 87-92.

Zhu, J., Nestor, K.E., Patterson, R.A., Jackwood, D.J. and Emmerson, D.A. (1996) Measurement of genetic parameters within and between turkey lines using DNA fingerprinting. *Poultry Science* **75**, 439-446.